Measuring what matters?

Exploring the use of values-based indicators in assessing Education for Sustainability

Ashley Jay Brockwell

Measuring what matters?

Exploring the use of values-based indicators in assessing Education for Sustainability

Ashley Jay Brockwell

Thesis

submitted in fulfillment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 28 May 2019

at 1100 in the Aula.

Acknowledgements

First of all, I would like to thank Marie Harder, founder and head of the Values and Sustainability Research Group at the University of Brighton and China Thousand Talents Professor at Fudan University, Shanghai, without whom this endeavour would never have been started. Marie, I'm enormously grateful to you, not only for believing in me and encouraging me to embark on a PhD, but also for your tireless energy and enthusiasm in leading the group in co-designing the overarching ESDinds and 'Starting from Values' projects, and the smaller research studies within it that provided the basis for these and other published works. This PhD has been a long and sometimes frustrating journey, and it's entirely thanks to you that I was able to persevere with it after the initial disappointments, enrol as an external candidate at Wageningen with you as a co-promotor, and finally – after more than seven years - see it through to completion.

Huge thanks are also due to my promotor Arjen Wals and co-promotor Peter Tamás for your tireless support – not only academic, but also practical, logistical and emotional. Thank you both for putting up with my succession of personal and professional crises, gently nudging me back to the thesis work every time I drifted away, helping me get over my inferiority complex and impostor syndrome, guiding me to the greatly-appreciated financial support from Wageningen University and Research at a critical time, and providing valuable guidance on the content, structure and tone of the introduction and discussion chapters. On that note, I am also grateful to the administrators at Wageningen School of Social Sciences for facilitating the enrolment process, especially in relation to my name change, and the stipend payment. I greatly appreciate the help provided by Erwin Hellegering with cover design and printing of the thesis.

Much of the research included in this thesis built on the earlier efforts of other researchers who designed and conducted field work, and created resource materials, during the ESDinds project. I would like to say a special thank you to my former colleagues in the Values and Sustainability Research Group at the University of Brighton - Elona Hoover, Firooz Firoozmand, Ismael Velasco, Tiina Jaatinen and Lee Stapleton – for your friendship, support and encouragement, as well as for the wonderful exchange of ideas and insights that helped to enrich and enhance these publications (and the many others that we created together). Prior to my involvement, the ESDinds project as we know it would never have existed without Dimity Podger, Poppy Villiers-Stuart, Arthur Dahl and the four partner civil society organizations: the Alliance of Religions and Conservation (ARC), Earth Charter International (ECI), ebbf and People's Theater.

Table of Contents

Abbreviations

The following abbreviations are used in this manuscript:

BSG-CSO	Research for the benefit of specific groups: civil society organizations
CSE	Comprehensive sexuality education
CSOs	Civil society organisations
ECI	Earth Charter Initiative
EFA	Education for All
EfSRL	Education for sustainable and responsible living
ESD	Education for sustainable development
ESDinds	a project funded by the European Commission's Seventh Framework Programme from 2008 to 2011 under the specialist funding scheme 'Research for the Benefit of Specific Groups Civil Society Organisations' (BSG-CSO) (see ESDinds, 2011)
FAO	United Nations Food and Agriculture Organization
GDP	Gross domestic product
HIV	Human immunodeficiency virus
IAEG-SDG	Inter-Agency and Expert Group on Sustainable Development Goal Indicators
IUD	Intrauterine device
LEDCs	Less economically developed countries
MDGs	Millennium Development Goals
QL	Qualitative
QN	Quantitative
RtD	Research through design
SD	Sustainable development
SDGs	Sustainable Development Goals
UNECE	United Nations Economic Commission for Europe
UNEP	United Nations Environment Program
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNICEF	United Nations International Children's Emergency Fund

List of Figures

List of Tables

Chapter 1:
Introduction

Chapter 1. Introduction

William Bruce Cameron, an American professor of sociology, wrote in 1963, 'Not everything that counts can be counted, and not everything that can be counted counts' (Cameron, 1963). This statement relates to one of the greatest paradoxes faced by educators, evaluators and policy-makers alike: namely, that the most highly-valued educational outcomes are often precisely those which are the most difficult to measure. In particular, the majority of educational assessments and evaluations - whether they focus on students, teachers, projects, institutions, or even entire national education systems – are not designed to measure the extent to which the education in question is helping (or, indeed, hindering) efforts to set human societies on a more sustainable course.

A similar sentiment was expressed, in a more pragmatic way, by the authors of the Millennium Development Report (United Nations, 2015a). In reviewing the successes and failures of the Millennium Development Goals and envisioning a new international development agenda for 2016-2030, they entitled one of their chapters 'Measure what we treasure: sustainable data for sustainable development' (p. 10). This was followed up with the subheading 'What gets measured gets done' (*ibid*) and, later in the section, 'Only by counting the uncounted can we reach the unreached' (p. 11). Yet the casual use of 'we' in these statements implies a unaniminity of thought, and a universality of human values, that is very far removed from reality – all the more now, in 2018, than when the report was first drafted. In a world where far-right extremism, white nationalism, systemic homophobia and transphobia and entrenched misogyny have all gained legitimacy at the highest levels, the naïve assumption that 'we' (as human beings) all treasure the same things has been revealed as untenable. In this light, the question of how the outcomes of education are assessed and evaluated[1] is more urgent than ever.

In this thesis, I propose a rethinking of educational assessment and evaluation at the level of their most fundamental question: *whose* values are taken into account (and, by implication, whose are ignored), and against whose criteria the respective individuals,

---

[1] Throughout this thesis, I use the term 'assessment' to refer to efforts to gauge educational outcomes at the level of individual learners and teachers, and 'evaluation' to refer to the overall examination of educational initiatives, institutions or national education systems. These are fundamentally intertwined, but they are not the same thing.

initiatives, institutions or systems are being measured (c.f. Springett, 2001). The method outlined here is based on two key principles: (i) *engaging multiple stakeholder groups* in defining which outcomes are the most valuable, meaningful and worthwhile within a specified context, rather than uncritically adopting the values and priorities of the dominant group or institution; and (ii) developing multi-level evaluation frameworks that take these diverse views into account, through a participatory process of choosing and prioritising indicators to be measured. This shifts the question from 'what *can* be measured, using current methods and datasets?' to 'what *should* be measured, to ensure progress towards the type of education that will ensure the survival of humanity and the Earth, into the 22nd century and beyond?'

1.1.   Key concepts and theories

The approach to assessment and evaluation design that I outline in this thesis can be described as *inductive*, in that the indicators are drawn out from participants' comments rather than being derived from a theoretical framework (deductive) (c.f. Fereday & Muir-Cochrane, 2006). It is also *intersubjective*, in the sense of relying on shared understandings created through what the German philosopher Martin Buber (1979) terms 'genuine dialogue' (in which each person acknowledges that the other has a separate existence and a different perspective, and strives to establish an authentic relationship with the other[2]) within a context of lived experience and practical activity (Talamo & Pozzi, 2011).

In both of these respects, my approach builds on well-established traditions of grassroots involvement in identifying questions, proposing solutions and collecting data - such as participatory evaluation, participatory design, co-design, adaptive management, transformative mixed methods research, action research, and 'research through design'. Yet the sense of separation that pervades academia, educational practice and professional practice, and the ever-increasing plethora of disciplines and specialities within each of these broad areas, has made it very challenging for specialists in these diverse areas to meet and

---

[2] Buber contrasts 'genuine dialogue' with two other forms of interaction, namely 'technical dialogue' where the focus is on gaining an objective understanding, e.g. giving or requesting instructions, and 'monologue disguised as dialogue', where

learn from each other. Meanwhile, the institutions tasked with defining ways to 'measure what we treasure' – from learners and teachers in schools, all the way up to the Inter-Agency Expert Group on Indicators for the Sustainable Development Goals (IAEG-SDG) – continue to pursue strategies that are neither inductive nor intersubjective. Instead of clearly negotiating 'best case scenarios' between diverse groups, and then using assessment and evaluation to track and advance progress towards them, their work tends to be based on the principles of tradition (what has historically been measured) and convenience (what is easy to measure).

There are, potentially, almost as many possible contexts for inductive and intersubjective design as there are different scenarios for evaluation or assessment. For the purposes of this thesis, I have chosen to focus on evaluation (rather than the assessment of individual learners) within the context of formal and non-formal education for sustainability, and specifically the evaluation of progress towards the 'Education for Sustainable Development' (ESD) target within the post-2015 Sustainable Development Goals (SDGs), at a variety of levels. While it clearly has substantial funding implications in comparison to today's convenience-based methods, the use of inductive and intersubjective approaches to indicator design within this context also raises the intriguing possibility of envisioning and creating radical shifts in local, national and even global sustainability discourses.

I will present and critically discuss one specific example of an emerging inductive and intersubjective approach to sustainability assessment within the context of the United Nations Sustainable Development Goals, and in particular, the Target on `Education for Sustainable Development' within Goal 4, the Sustainable Development Goal relating to quality education.   This particular example, namely an approach to the creation of 'values-based indicators' that has been referred to as *WeValue*, emerged within a particular historical context – namely, the European Union's interest, at the beginning of this decade, in funding what it termed 'research for the benefit of civil society organisations' (BSG-CSO) through its Seventh Framework Programme. The BSG-CSO scheme sought to challenge and subvert the established process of academic knowledge creation, in which questions are formulated by university-based scholars and answered by doing research *on* civil society, by establishing international consortia of CSOs and academics to co-define both the questions and the means of answering them (European Commission, 2007). During my employment in the Values and Sustainability Research Group at the University of Brighton from 2010-2016 inclusive, I was a

member of the consortium (led by Professor Marie Harder at the University of Brighton, and incorporating representatives of the University of Brighton, Charles University Prague, and four civil society organisations) that was tasked with co-designing and co-delivering one of these projects. The project itself was entitled 'ESDinds: The Design of Values-Based Indicators and Assessment Tools for Civil Society Organisations Promoting Education for Sustainable Development' (ESDinds, 2011). It was followed up with another collaborative multi-stakeholder project led by Professor Marie Harder, this time with co-investigators from The Open University (Theodore Zamenopoulos and Katerina Alexiou), the University of Exeter (Julian Brigstocke), and the civil society organisations The Glass-House (Sophia de Sousa), Fossbox (Paula Graham), Silent Cities (Justine Gaubert), and Blackwood Foundation (Colin Foskett). The 'Starting from Values: Evaluating Intangible Legacies' project was funded by the Arts and Humanities Research Council to explore the legacies of earlier Connected Communities collaborative research projects (Hoover, Burford, Dredge, & Harder, 2015).

The published works presented in this thesis are those for which I was directly responsible for identifying the research questions, carrying out the main literature review, and conducting the data collection (or compilation of existing datasets) and analysis. Other colleagues named as co-authors in the published papers were responsible for the overall design of the large-scale ESDinds research project, and contributed in varying degrees to the initial design of these separate studies within it, including the formulation of research questions; the identification of additional references for the respective literature review sections[3]; and/or the initial creation of datasets which I subsequently compiled and analysed (Chapters 2 and 4 only). My co-authors, along with my promotors and the peer-reviewers nominated by the respective editorial boards, also acted as 'critical friends' throughout the process of drafting and submitting the papers.

These four publications represent only a portion of my research output between 2012 and 2016, with the remainder being incorporated into publications with more substantial co-authorship contributions from colleagues (e.g. (Burford, Hoover, Jarvis, & Harder, 2014; Burford, Velasco, et al., 2013; Harder, Burford, & Hoover, 2013; Harder, Velasco, et al., 2014;

---

[3] Specifically, Elona Hoover identified the business ethics references in Chapter 2 (which was initially submitted to *Organisation Studies* in 2014) in the process of revising and re-contextualising it for submission to the *Journal of Business Ethics.* I was responsible for revising the paper a second time for submission to the online journal *Sustainability,* in which it was finally published in 2016.

Hoover et al., 2015; Podger, Hoover, Burford, Hak, & Harder, 2016; Podger, Velasco, Amezcua Luna, Burford, & Harder, 2013; Ribeiro et al., 2016). They describe, respectively, the first successful trial of an inductive and intersubjective approach to evaluation in CSOs providing non-formal education (Burford, Hoover, Stapleton, & Harder, 2016: see Chapter 2)[4]; the theoretical justification for applying such approaches to the Sustainable Development Goals (Burford, Hoover, et al., 2013: see Chapter 3); a direct comparison of an inductive/ intersubjective ('values-based') approach with the approach that was actually adopted by the IAEG-SDG in identifying indicators for the Sustainable Development Goal pertaining to education for sustainable development (Burford, Tamas, & Harder, 2016: see Chapter 4); and an exploration of applicability to formal education, including (Burford, Hoover, Dahl, & Harder, 2015: see Chapter 5);.

The core of all these publications can be broadly summarised as a five-step process of evaluation design. In the first two steps, participants are (i) encouraged to articulate what matters to them in their own words (usually through a combination of reflection on past experiences and projection of `best-case scenarios' for the future: c.f. Sanders and Stappers, 2012); and then (ii) challenged to rethink these initial assumptions by responding to a set of stimulus materials. These materials may take the form of 'proto-indicators', actual indicator sets or assessment tools created by other groups, or 'trigger statements' that provide prompts for discussion. These two phases, which I have respectively referred to as the *elicitation* and *challenge* phases[5] in this thesis, are followed up with collaborative processes of (iii) prioritising indicators and finalising their wording (the *consolidation* phase); (iv) developing context-appropriate assessment tools for the respective indicators; and (v) collecting and analysing the evaluation data (see also Chapter 5, Section 5.1.3).

While it is by no means the *only* possible way of initiating inductive and intersubjective processes of evaluation design, this five-step strategy – which evolved iteratively during the course of the ESDinds project and the subsequent 'Starting from Values: Evaluating Intangible Legacies' project – is conducive to 'genuine dialogue' as defined by Buber (1979), and

---

[4] The works submitted for this thesis were published under the name of Gemma Burford, which I used for all purposes before the legal adoption of my current name, Ashley Jay Brockwell, by deed poll in March 2018.
[5] The phrase 'elicitation phase' was used by the research team during the ESDinds project, but 'challenge phase' and 'consolidation phase' are phrases that I have coined retrospectively as a result of reflection on the process. The significance of the challenge phase is discussed further in Chapter 6.

intersubjectivity as defined by Talamo and Pozzi (2011) in that it first gives participants space to clarify their own standpoints, and then provokes them to engage more deeply with the views of other stakeholders, who may or may not have a place at the table in their own right. The inclusion of a 'challenge' phase can give people confidence to disrupt and rethink some of their convenient or culturally-validated assumptions, probe into their own motivations for valuing particular outcomes, or even identify ways in which their understanding of 'needs' (and of what matters in life) may have historically been manipulated, e.g. by corporate marketing or by the views of authority figures.

The work described in this thesis opens up intriguing possibilities for innovation in evaluation design, both in relation to monitoring progress towards the Sustainable Development Goals, and in relation to helping teachers, learners and policy-makers to understand the impact of educational initiatives (whether formal or non-formal) in deeper and more nuanced ways. It is also directly applicable to an emerging area of educational research and praxis known as 'T-learning', which was not widely known at the time when it was published. 'T-learning' is a term used to describe modes of learning that are both *transformative*, in the sense of working towards social, environmental and economic justice, and *transgressive,* in the sense of promoting systemic decolonisation and overturning the well-established assumptions that underpin unsustainable worldviews and practices (Lotz-Sisitka et al., 2016; Lotz-Sisitka, Wals, Kronlid, & McGarry, 2015). To these I have added a third T: *transdisciplinary*, not only in the sense of blurring the boundaries between academic disciplines and professional practice traditions, but also in the theoretical sense described by Basarab Nicolescu (2002) of deliberately constructing an 'included middle' and applying a 'both/and' logic to reconcile seemingly discrete bodies of knowledge and practice.

As I will discuss further in Chapter 6, I have coined the phrase 'triple-T learning' to refer to learning initiatives that are simultaneously transformative, transgressive, and transdisciplinary. Owing to their ability to promote meta-cognition ('thinking about thinking') and meta-learning ('learning about learning'), inductive and intersubjective approaches to assessment and evaluation are a perfect match for such 'triple-T' learning initiatives. They exemplify not merely 'education *for* sustainability', but also 'education *as* sustainability' (c.f. Sterling, 2001). Seen through a competency lens, inductive and intersubjective assessment (or evaluation) can contribute to building the very competencies that they seek to measure,

such as the ability to listen with compassion, understand and celebrate cultural diversity, or respond critically to different viewpoints.

An important point to note, which I discuss briefly in Section 3.4.2 of this thesis and revisit in Chapter 6, is that there are many limitations and caveats associated with inductive and intersubjective approaches to designing assessments and evaluations - especially in today's challenging global political context. It is important to establish guidelines or codes of conduct to minimise the risk of misuse (or deliberate abuse) of these approaches, potentially leading to perverse effects. Without due attention to human diversity, and beyond that, to the multi-subjectivity of the more-than-human world (as discussed, for example, by ecophilosopher David Abrams (1996), there is a risk that approaches might reify anthropocentric and instrumentalist understandings of 'nature' in the sense of 'natural resources' and contribute to the further silencing of marginalised social groups. These include, among many others, Indigenous communities that are already struggling to defend non-anthropocentric and anti-instrumentalist positions.

The parallel emergence of triple-T learning initiatives, on the one hand, and of inductive and intersubjective approaches to their evaluation, on the other, could offer a potential mechanism for disrupting deeply-entrenched patterns of oppression of Indigenous and local communities and the natural environments that they steward. However, this will happen only if educators and leaders are willing to make a long-term commitment to decolonising education - ensuring that the transgressive aspects of triple-T learning are not downplayed or dodged for the sake of political expediency, and genuinely valuing marginalised perspectives rather than paying lip service to them.

1.1.1. To 'values-based indicators'…and beyond

The term 'values' can serve as a helpful shorthand for those behaviours, attitudes, etc. that are individually or collectively viewed as valuable, worthwhile and meaningful within a specified practical context. As such, it provided a starting point for much of the work documented in this thesis. In exploring the potential for developing intersubjective and inductive approaches to educational assessment, especially in Education for Sustainability, the ESDinds project originally framed its research questions in terms of developing and testing

'values-based indicators and assessment tools' (ESDinds, 2011). Specifically, we were starting from a particular understanding of values as "principles or standards that guide behaviour", i.e. *ethical* values that carry an imperative for action, as opposed to merely "judgements about what is important in life" (c.f. Oxford English Dictionary, 2013), as I discuss in some depth in Paper 2.

The specific context of indicator development implies measurement and assessment, and attempts to 'measure values' with reference to predetermined constructs have already been recognised as problematic, especially in cross-cultural contexts (Braithwaite & Law, 1985; D. Brown & Crace, 1996; Peng, Nisbett, & Wong, 1997). While individual and organisational values measurement tools developed by social psychology researchers such as Milton Rokeach (1973, 1979b) and Shalom Schwartz (1992, 1994; 1987; 1990; 1990; 1999) have long been, and continue to be, used for a wide variety of research and education purposes, they impose rigid values frameworks that are predefined by researchers and purport to be universal across all cultures and political affiliations. As well as being challenged by other researchers (Bardi, Lee, Hofmann-Towfigh, & Soutar, 2009; Braithwaite & Law, 1985; D. Brown & Crace, 1996; Peng et al., 1997), these frameworks and their underlying assumptions are in conflict with contemporary constructivist understandings of values as negotiated, dynamic, fluid, 'intangible', and fundamentally situated within very particular geographical, political and cultural contexts. In Section 1.2, *Scientific Contribution,* I critique some of these established approaches in the field of values measurement and explain why a fresh approach is needed.

If we were to follow the 'intangibility-of-values' argument to its logical conclusion, it would imply that efforts to measure what matters to people in the context of Education for Sustainability are inevitably doomed to failure - because the entire concept of 'measurement' is predicated on an objective definition and a fixed, unchanging set of criteria or benchmarks. This argument is thus highly problematic in the light of the Millennium Development Report's observation that processes of measurement not only reflect, but also contribute to *defining,* what is important to an organization or a society. As the report states, 'what gets measured gets done' (United Nations, 2015a, p. 10) and by implication, anything that is left unmeasured is all too often neglected. This dilemma hints at a fundamental problem at the heart of policy-making: namely that national and global agendas are largely shaped not by what matters most

to diverse constituent groups, but by what can be most conveniently assessed and reported using readily available data collection instruments (Bell & Morse, 2011). If there is a general consensus that the values that matter most to a certain stakeholder group are not 'measurable', they may be overlooked with impunity by policy-makers - thereby contributing to the further marginalisation of that group and the imposition of dominant economic and political narratives.

I will develop some of these arguments in Section 1.3, *Societal Relevance,* in which I also highlight some recent concerns relating to measurement and evaluation in Education for Sustainability within both formal and non-formal settings – and in particular, the emergence of 'triple-T learning'. I demonstrate the inadequacy of conventional 'top-down' approaches to evaluation and assessment, and highlight the urgent need for innovative approaches that are congruent with the triple-T model.

The approach that I propose in this thesis is grounded in an iterative and praxis-oriented methodology. It closely resembles the 'Research through Design' approach that was initially proposed in the 1990s by Christopher Frayling and Bruce Archer at the Royal Society of Art, who framed it respectively as 'research through art and design' (Frayling, 1993) and 'research through practitioner action' (Archer, 1995), and subsequently developed by researchers in both architecture and human-computer interaction (e.g. Forlizzi, 2014; Zimmerman, Forlizzi, & Evenson, 2007). My early case studies (e.g. Burford et al., 2012) were not initially framed in relation to the Research through Design (RtD) literature, I have since found this literature to yield a set of principles that can usefully be applied to the general context of inductive and intersubjective indicator design as well as to the specific case studies outlined below (see Section 1.3, *Methodological Perspective and Methods*).

An important point to note about the worked example of inductive and intersubjective indicator design presented in this paper is that it was initially designed as a tool for evaluators (and, to a lesser extent, program managers) within the ESD arena (Burford, Velasco, et al., 2013; ESDinds, 2011; Podger et al., 2010; Podger et al., 2013). As such, it has a firm grounding in the academic literature pertaining to project evaluation – especially in relation to well-established concepts of process-based, participatory and utilization-focused evaluation, and the 'process use' of evaluation - i.e. benefits derived from taking part in an evaluation process, which are independent of the evaluation findings (Cousins, 2007).. While space does not

permit me to explore these concepts in detail here, I have discussed them at length in a literature review that was published as part of an earlier co-authored paper (Burford, Velasco, et al., 2013). The reader is encouraged to refer to this review for a clearer understanding of these key concepts and theories, and of how my work is positioned within the academic field of evaluation studies. Likewise, for a deeper understanding of the concepts of 'validity' and 'participation' as they pertain to the four publications included in this thesis, I refer the reader to my literature reviews published in co-authored papers in the *Journal of Environmental Management* (Harder, Velasco, et al., 2014) and *Design Issues* (Harder et al., 2013) respectively.

Following the introductory chapter, the main body of the thesis will consist of three articles that I have published in the international peer-reviewed scientific journal *Sustainability* and one chapter published in an edited volume, organised as follows:

> **Chapter 2:** Burford, G., *et al.* (2016a). An unexpected means of embedding ethics in organizations: preliminary findings from values-based evaluations. *Sustainability* **8**(7), 612; doi:10.3390/su8070612
>
> **Chapter 3:** Burford, G., *et al.* (2013). Bringing the 'missing pillar' into Sustainable Development Goals: towards intersubjective values-based indicators. *Sustainability* **5**(7), 3035-3059; https://doi.org/10.3390/su5073035
>
> **Chapter 4:** Burford, G., *et al.* (2016b). Can we improve indicator design for complex Sustainable Development Goals? A comparison of a values-based and conventional approach. *Sustainability* **8**(9), 861; https://doi.org/10.3390/su8090861
>
> **Chapter 5:** Burford, G. *et al.* (2015). Making the invisible visible: designing values-based indicators and tools for identifying and closing 'value-action gaps'. In R. J. Didham, D. Doyle, J. Klein and V. W. Thoresen (eds.) *Responsible Living: Concepts, Education, and Future Perspectives.* Berlin, Heidelberg: Springer Verlag, pp. 113-134.

The titles of these works have been revised for the purpose of their inclusion as chapters in this thesis, to highlight the coherence between them and the extent to which they build on one another.

Following the presentation of the preliminary findings illustrating organisational benefits from values-focused evaluation in non-formal 'education for sustainability' contexts (Chapter 2), I first explore the theoretical feasibility (Chapter 3) and then provide a practical

example (Chapter 4) of the application of this work to the context of the Sustainable Development Goals. I then illustrate how the approach can be adapted for formal education contexts, not only as a method of designing novel assessments of students' learning outcomes (which I do not discuss explicitly in this thesis, but have covered elsewhere – see Burford, Hoover and Harder, 2015), but also as a tool for stimulating reflection to increase the coherence between values, discourse and action (Chapter 5).

In Chapter 6, I examine these published works with the benefit of hindsight. I distinguish the limitations and challenges that are inherent in the creation of inductive/ intersubjective approaches to indicator design from those that are associated with specific projects or activities, and constitute artefacts of their history. Specifically, I critique my earlier work in the light of questions such as 'what is problematic or invalid?', 'what was missed or overlooked?', 'what was not fully appreciated as a positive?' and 'what has changed in the landscape since these papers were published?', while also highlighting key areas of theory - notably Basarab Nicolescu's three axioms of transdisciplinarity (Nicolescu, 2002) and David Abrams' notion of the multisubjectivity of the more-than-human world (Abram, 1996) - whose relevance to the work was initially underestimated or insufficiently developed.   In the final section, I propose a way forward that seeks to mitigate some of the risks of this methodological innovation by using Research through Design as a strategy to cultivate grassroots artivism and 'triple-T learning'.

1.2. Scientific Contribution: Addressing the 'Values Problem'

The published research presented in this thesis contributes to the scholarly debate about how people's diverse views on what is meaningful, worthwhile and valuable within a specific project context can be used as a foundation for assessment and evaluation. I do this by critiquing the theories and concepts that have historically dominated conversations about how to measure what matters to people, and introducing my own explorations in Research through Design as an alternative framing of the issue.

In social psychology, the tradition of positivist ontologies of values - which takes as its underlying assumption the notion that human values are organised as discrete, universal and

consistent types, forming a conceptual structure that is independent of culture and politics - dates back more than a century to the work of Eduard Spranger (1882-1963). In his book *Lebensformen,* first published in German in 1914 and translated into English as *Types of Men* (*sic*) by P. J. W. Pigors (1928), Spranger distinguishes six 'ideal types' of value attitudes, which he lists as *theoretical* (primarily concerned with the discovery of truth), *economic* or utilitarian (passionate about gaining a return on investments), *political* or individualistic (concerned with achieving power and using it to influence others), *aesthetic* (dedicated to form, harmony and self-actualisation), *social* (committed to helping people achieve their potential), and *religious* or traditional (pursuing a divine ideal or higher meaning in life). This work provided the theoretical foundation for the first popular values measurement tool, the Study of Values (SOV) questionnaire (Allport, Vernon, & Lindzey, 1951). This personality measure was widely used in counselling, pedagogy and research from the 1930s to the mid-1970s (Kopelman, Rovenpor, & Guan, 2003) and required respondents to state how they would behave in each of 55 hypothetical scenarios, choosing from two or four possible answers in each case. On the basis of their responses, they were assigned a numerical score for each of the six value types. An updated version of this tool has been created by Kopelman et al. (2003), without changing the underlying model.

Milton Rokeach (1973, 1979a, 1979b, 1985), widely acknowledged as a pioneer of the study of values in social psychology, echoed Spranger's realist assumptions in assuming that there are a limited number of value types and that they are all universally recognized. Cross-cultural differences, according to Rokeach's theory, are attributable only to differences in the priority order of the values. Rokeach expanded the list of value types from six to 36, dividing them into two higher-order categories that were each represented by 18 values. These are, respectively, `instrumental values', which are expressed in the form of adjectives and correspond to modes of conduct that are allegedly viewed as socially desirable across all cultures (e.g. *courageous*, *loyal*, *helpful* or *capable*), and `terminal *values'*, which are expressed in the form of nouns and represent idealized end-states of existence (e.g. *equality, an exciting life, mature love,* or *a world of beauty*). The Rokeach Values Survey (Rokeach, 1979b) presents these items in alphabetical order and requires the respondent to rank them in order of importance as guiding principles in their livesA modified version of this survey, in which ranking is replaced by the use of a rating scale, was proposed by Braithwaite and Law

(1985); see also Debats (1996). The Schwartz Value Survey (SVS) (S.H. Schwartz, 1992, 1994; S. H. Schwartz & Bilsky, 1987; S. H. Schwartz & W. Bilsky, 1990; S.H. Schwartz & W. Bilsky, 1990) also advocates a rating method, this time based on a standard list of 56 'values'.

A problem shared by the Rokeach and Schwartz surveys is that the items being ranked or rated are 'value-words', rather than predefined value constructs. In effect, the observable indicator for the values held by an individual is the set of rankings or ratings that they assign to the respective value-words, *according to their own personal understanding of these words*. There have been no systematic attempts to examine cross-cultural differences in the definition and framing of value-words (Peng et al., 1997). We also have no information about the extent to which users support the researchers' perception of the 36 value-words of the RVS, or the 56 value-words of the SVS, as representing discrete and clearly bounded constructs. It might be anticipated, for example, that some individuals would perceive a certain overlap between their concepts of a person who is *capable,* one who is *independent* and one who is *responsible* (instrumental values from the Rokeach Value Survey: Rokeach, 1973).

Schlater and Sontag (1994) have put forward a more fundamental objection to this abstract level of analysis: as values function as criteria for making judgements about 'objects' in the broad sense (not only inanimate objects, but also events, ideas, states and people), the interaction between subject and object is of primary concern. Omitting any mention of the *object* on which values focus, and discussing them in a contextual vacuum, may thus lead to unwarranted and misleading conclusions. Citing earlier work by Schlater (1969) in which users were asked the question "what should be done, and why?" in relation to projective stories about family decisions, they argue for an 'open' approach to measuring personal values, i.e. one in which permissible responses are not predetermined. More recently, Inglehart and Welzel (2005) have developed a 259-item questionnaire, combining ranking, rating and forced-choice items, as the basis of their 'World Values Survey' conducted in over 100 countries. In these open models, operationalization, tool development and data collection all precede conceptualization. In practice, this means that researchers use either a general prompt (Schlater, 1969) or a very broad spectrum of indicators (Inglehart & Welzel, 2005) to elicit generic values content from individuals. However, discrete value categories

are still imposed by the researcher(s) – in this case, at the data analysis stage – without any attempt at securing validation from the respondents.

A more participatory approach has been demonstrated by Reino and Vadi (2010) in their construction of an 'Organizational Values Questionnaire' (OVQ). This entailed brainstorming a list of keywords for each organizational value type, creating statements for each keyword, consulting 'independent experts' about which items best fitted which keyword(s), rejecting those with only moderate inter-rater reliability, constructing and testing new items, and presenting the draft questionnaire to a new expert group for validation. This illustrates a move away from researcher-led subjectivity in the assignment of value categories, and towards *intersubjectivity* – the creation of shared understandings through dialogue, rooted in a shared context of practical activity.

As I discuss in Paper 2, intersubjectivity can provide a way to overcome the impasse between avoiding rigid theoretical constructs on the one hand, and dismissing the whole idea of measurement on the other. However, the intersubjectivity described by Reino and Vadi (2010) is still very limited in its scope, limiting participation to arbitrarily selected experts rather than the stakeholders themselves. As such, it would be insufficient to meet the evaluation challenges posed by 'triple-T' learning, which seeks to cross or dissolve disciplinary boundaries (transdisciplinary), disrupt or invert well-established paradigms and power structures (transgressive), and create social change by giving agency to the marginalized and silenced (transformative). The papers that form the body of this thesis present some initial explorations in inductive and intersubjective indicator development that may be compatible with triple-T learning, while the discussion chapter draws out some of the limitations of this work and sets out a research agenda for the future.

1.3. Societal Contribution: How Can We Measure What Different People Treasure?

In outlining the societal relevance of this work, I will focus on three areas in which inductive and intersubjective approaches to assessment design and/or evaluation design can make original contributions to policy and practice:

(i) **The values foundations of the shift from 'Millennium Development Goals' to 'Sustainable Development Goals'.** Among the implications of this transition is the *need for innovative approaches to monitoring and evaluating progress towards the realization of SDGs, at the 'big-picture' level*. The application of a values-centred monitoring and evaluation approach is thus proposed as an effective way to address the challenge of multiple stakeholders with competing agendas. However, it needs to be carefully managed to mitigate the risk of escalating tensions by bringing hidden value conflicts to the surface.

(ii) **The central role of education in general, and what has been referred to as 'Education for Sustainable Development' (ESD) in particular, within the overall landscape of the Sustainable Development Goals (SDGs).** This flags up the *need for high-level monitoring and evaluation of the `ESD' target as an essential foundation for the achievement of other targets,* in the sense that very little can be achieved within the SDG framework in the absence of appropriate education. However, there is also a need to problematize the concept of 'ESD' as a term that is both derived from, and intimately connected to, contested discourses of 'sustainable development' and of 'development' in the wider sense.

(iii) **The emergent shift in thinking about education for sustainable and responsible living,** from its early days when it was primarily identified with `environmental education' (in the sense of educating people about environmental challenges and their likely impact), to more complex approaches which focus on agency and the development of competencies, and most recently the concept of transformative and transgressive learning (T-learning) as discussed in Section 1.1 above. Set within the well-established discourse of accountability and results-based management in education (c.f. UNESCO, 2017), this thesis highlights the *need for indicators and tools that can be utilised in real-world educational settings* for three interrelated purposes: (a) to identify valued learning and practice outcomes at the individual, group and institution levels, especially those that might be concerned more with the development of competency, agency and the willingness to 'think outside the box' than with the acquisition of factual knowledge about

environmental problems; (b) to monitor progress towards these outcomes; and (c) to create or enhance teaching and learning activities that can contribute towards the achievement of these desired outcomes.

1.3.1. From 'MDGs' to 'SDGs': a global paradigm shift?

The Millennium Declaration (United Nations General Assembly, 2000) set out eight specific global development goals to be achieved by 2015. These were to eradicate extreme poverty (controversially defined as "living on less than a dollar a day") and hunger; to ensure that all boys and girls complete a full course of primary education; to eliminate gender disparity at all levels of education; to reduce the child mortality rate among under-fives by two-thirds; to reduce the maternal mortality rate by three-quarters; to halt and begin to reverse the spread of AIDS and the incidence of malaria and other major diseases; to develop a global partnership for development; and to 'ensure environmental sustainability'. Achieving environmental sustainability, in turn, required member states to "integrate the principles of sustainable development into country policies and programmes", reverse the loss of environmental resources, halve the proportion of people lacking access to safe drinking water, and "achieve significant improvement" in the lives of slum dwellers.

The Millennium Development Goals were generally understood to be relevant only to less economically developed countries (LEDCs) and the process of their creation was, as noted by Waage et al. (2010), grounded in development trends and fashionable campaigns of the 1980s and 1990s rather than a systematic global analysis of priority needs. They were donor-driven, created with minimal involvement of LEDCs, and rooted in an unquestioning acceptance of GDP growth as the overarching mechanism for poverty reduction (United Nations, 2015a; Waage et al., 2010); although, as noted by Manning (2009), the neoliberal foundation of the goals was 'tempered' by the inclusion of human welfare goals such as health and education, to improve buy-in from developing country governments and particularly from civil society. They arose in a context where 'results-based management' was becoming the norm, in international development as in the corporate sector, and narrowly focused numerical targets were pursued at the expense of an integrated, long-term approach (Ebrahim, 2002, 2003). The MDGs have also been criticised for being unachievable and

simplistic, dismissing national needs, failing to specify accountable parties, and reinforcing a 'vertical intervention' model (Fehling, Nelson, & Venkatapuram, 2013). Baseline data for many of the MDG indicators were never collected, making it impossible to determine whether or not the targets had been met; and some indicators, such as malaria-specific mortality in malaria-endemic countries, were dismissed even by the UN's own scientific experts as essentially immeasurable (Attaran, 2005). This resulted in a strong focus on those goals that were easier to implement and measure (Lomazzi, Borisch, & Laaser, 2014). To engage in a systematic critique of this particular paradigm of international development is beyond the scope of this thesis, but a key point to note is that, as highlighted by Ferguson (1994), development discourse typically renders politics invisible and misrepresents political concerns as technical problems to be solved. As such, the top-down approach of the MDGs provides a stark contrast to the inductive and intersubjective approaches to indicator design (and data collection) that are championed in this thesis.

Perhaps unsurprisingly, the official report on the MDGs revealed that none of the goals were achieved by 2015, although all countries had "made progress on at least one target" (United Nations, 2015, p. 62). Low buy-in, especially from LEDC governments and global civil society, was named as a key challenge, deriving largely from the fact that the goals were very narrowly focused and did not always correspond well to pre-existing national priorities. The focus on priority diseases at the expense of structural improvements in health care was particularly problematic. The report also admitted that "the MDG focus on outcomes such as poverty reduction without particular attention to the underlying causes has led in some cases to undesirable, unintended and often unsustainable consequences" (p. 63). The challenges of evaluation were specifically highlighted in the MDG Report, which demanded a 'data revolution to improve the availability, quality, timeliness and disaggregation of data' to support the emerging sustainable development agenda (p. 10), and described this revolution as 'a joint responsibility of governments, international and regional organizations, the private sector and civil society' (p. 13).

In 2012, the United Nations initiated a global process of consultation with a view to creating 'Sustainable Development Goals' (SDGs) to succeed the MDGs. The intent was that these goals would no longer be viewed as primarily something imposed by donors on LEDCs, but as an agenda for human and planetary survival that is applicable – at least in principle –

to all Member States alike (United Nations, 2015b; United Nations Development Group, 2013; United Nations Millennium Campaign, 2015).   This involved a 'global survey entitled 'My World' that engaged over 8.5 million people in reflecting on possible goals and prioritising them, albeit from a predetermined list, as well as 88 national consultations and 11 thematic consultations conducted by UN agencies to gather more detailed data (United Nations Millennium Campaign, 2015).   The result of the consultation was the creation of 17 SDGs, encompassing a total of 169 targets.

Sustainable development was initially defined by the World Commission on Environment and Development (1987, p. 54)as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs". This ambiguous definition fails to address the reality that different stakeholders have different, and often competing, needs – and indeed, that people's understanding of 'the needs of the present' is highly susceptible to manipulation by corporate interests, a principle that underlies the whole of the advertising and marketing industries. In particular, although sustainable development is widely recognised as encompassing environmental, social and economic dimensions, it usually focuses on *human* needs and takes an instrumentalist view of non-human species, in contrast to an ecocentric perspective that would value global and local ecosystems for their own sake and consider their needs at the whole-system level (Thompson & Barton, 1994).

In the absence of mechanisms for discussing its meaning, the term 'sustainable development' is often interpreted as referring to the achievement of long-term GDP growth, and many national policies have been built on this foundation (Bell & Morse, 2011). Specifically, in the SDGs an increase in GDP is named explicitly as one of the goals (SDG 8), while in the MDGs it was only implicit. This creates an inherent tension within the SDG framework, in that GDP cannot be increased indefinitely without exhausting the planet's natural resource base and compromising the achievement of the environmentally-oriented goals. The use of material accumulation as a proxy for wellbeing that is exemplified by GDP, and illustrates what Jickling (2013) describes as 'normalizing catastrophe', contrasts sharply with other ways of measuring human progress, such as the Happy Planet Index (Marks, 2006) and Kate Raworth's 'Doughnut Economics' (Raworth, 2017), which balance the need for human development with the imperative of minimising 'ecological footprints' and remaining within planetary boundaries. This calls into question whether the SDGs really constitute a

new paradigm of global development, characterised by greater inclusivity and wider participation, or if the process has been designed to create an illusion of diverse stakeholder engagement while leaving the underlying foundation of neoliberal values unchallenged since the MDG days.

These contradictions and tensions within the SDG framework are brought into focus by the question of how to monitor and evaluate progress towards the SDGs. Indicator development is, in itself, a highly politicised process often disguised as a technical challenge, in that the decision to evaluate a specific outcome inevitably has the consequence of directing human and financial resources towards its achievement at the expense of potentially competing outcomes. This can have unintended and often far-reaching implications (Bell and Morse, 2011). Insisting on a single set of SDG targets and a single set of indicators to be applied to all global, national, regional and local initiatives, in all circumstances, both denies the importance of diversity and increases the risk that goal achievement will be compromised by poor buy-in.

In response to this criticism, modern indicator frameworks are often characterised by the uncritical and unsystematic *accumulation* of indicators from different sources (Grainger, 2012). The 'conventional' SDG indicator development process that I describe in Paper 4 provides an excellent example of this phenomenon. While it may capture some of the diversity, and provide some space for contextual sensitivity, the accumulation of long 'shopping lists' of indicators can lead to fragmented and reductionist thought on the one hand, and invite a managerialist approach to accountability that is essentially unsustainable, on the other. It also provides a cover for educational initiatives that are successfully 'ticking the boxes' on uncontroversial indicators, while making no meaningful progress on less convenient or politically expedient outcomes.

In this thesis, I explore multiple perspectives on the core elements of sustainability, identifying aspects that are often overlooked in the traditional 'three-pillar' model (environmental, social and economic) - including cultural/aesthetic, political/institutional, and religious/spiritual dimensions. I illustrate that these 'missing' aspects may be seen as having common ground within the arena of ethical values, and that the creation of values-based indicators through the 'WeValue' process potentially allows some of these less tangible dimensions (which constitute an essential *foundation* for the achievement of the

36

environmental, social and economic outcomes)to be brought into the realm of monitoring and evaluation. This opens a space for a broader debate, in which – in a best-case scenario – diverse groups of people are given the agency to interpret the SDGs in relation to their *own* values, world-views and priorities, in parallel with the standard interpretations and indicators, and to create and evaluate programs that advance their own particular agendas while still contributing to the achievement of the broad SDG target. Such efforts may help counter the tendency for the SDGs, created through the widest and most inclusive global consultation process to date - to evolve into yet another mechanism for expanding bureaucratic power through a homogenising form of globalisation that pursues the 'measurable' at the expense of what is truly important for human and planetary survival   They may also mitigate the risk of civil society organisations disengaging from the SDG process and thereby compromising the possibility of goal achievement, as in the case of the earlier MDGs.

Inductive and intersubjective approaches to assessment and evaluation design are not without risk, and it is important to ensure that they are not manipulated to exacerbate existing tensions, or to expand bureaucratic power by reinforcing entrenched power dynamics. While the 'values challenge' phase can offer a potential safeguard against misuse and abuse, the design of the stimulus materials used in this phase can make the difference between a process that merely reifies and justifies unsustainable norms of excessive material consumption, and one that genuinely takes steps to overturn them. Incorporation of a wide range of values and perspectives – particularly from the most marginalised communities, and those with novel or radical views on what 'sustainability' might require – into the stimulus materials is crucial to the success or failure of the whole enterprise.   It is in the co-design of these materials that the principles of 'triple-T' (transformative, transgressive and transdisciplinary) learning are fundamentally important.

1.3.2. The central importance of education in the SDG framework

In addition to exploring the diversity of productivities of inductive and intersubjective approaches to assessment and evaluation design within the context of increasing stakeholder engagement with the SDGs in a generic sense, this thesis will also make the case that such approaches can be particularly useful within the specific field of education.

A study by Vladimirova and Le Blanc (2016), initially published as a working paper by the UN Department of Economic and Social Affairs, set out to investigate the linkages between education and the other Sustainable Development Goals. The authors conducted a content analysis of 37 flagship United Nations reports, which were assumed to provide a good indication of the 'science-policy interface', i.e. the key messages conveyed by the respective United Nations agencies to policy-makers. They found that the reports made direct reference to links between education and all the other SDGs, with the exception of SDG14 on oceans; and that the links were, in nearly all cases, bidirectional – exploring both education's influence on the other goal areas, and vice-versa. This means that the SDGs can be organised in the form of a conceptual map with education at the centre[6]. The links from education to the other goals are shown to operate through five main 'channels', namely (i) education increases labour productivity; (ii) it promotes entrepreneurship and increases people's capacity to innovate, (iii) awareness-raising can impact behaviour, e.g. with respect to climate change or sustainable consumption and production; (iv) education and training allow people to acquire specialised skills, e.g. in ecosystem management; and (v), most crucially in terms of this thesis, "education transmits values, changes social norms and enables empowerment" (ibid, p. 22). In the other direction, progress towards the other SDG goals can result in changed demands on the educational system, especially in terms of content; while a lack of progress in other goal areas may perpetuate barriers to education, or adversely impact educational outcomes.

The central position of education within the SDG landscape makes it ever more crucial to broaden the way in which concepts such as 'accountability' and 'quality assurance' are defined and played out in formal educational practice. This is because every goal and target has underlying values and assumptions, which may or may not be explicitly acknowledged. Pursuing education with a view to achieving the GDP goal in particular is liable to result in educational initiatives, both at the global level and at very local levels, which are incompatible with environmental protection and are likely to compromise the achievement of *all* the other SDGs. The argument that I make in this thesis is that inductive and intersubjective processes of assessment and evaluation design, and the resulting indicator sets and assessment tools, may constitute useful contributions to this debate. One reason is that they can *provide a*

---

[6] Another way of thinking about the SDGs is to view them in terms of 15 content-oriented goals, broadly concerned with 'people, planet and prosperity', and two process-oriented goals, concerned respectively with education (goal 4) and partnerships (goal 17) (Arjen Wals, personal communication, 2018).

means through which the values underlying different educational policies, systems, institutions and initiatives can be made explicit and, where necessary, challenged and transformed, especially in relation to their congruence with environmental, social, cultural and spiritual sustainability. The second reason is that these indicators and processes can *allow us to problematize and evaluate the assumed connection between 'education for sustainable development' and 'sustainable development'*, respectively, which cannot always be taken for granted.

The rise of 'accountability' as a dominant discourse in education is discussed in depth by Suspitsyna (2010), who relates it to Foucault's concept of governmentality as the interface between techniques of state control (also referred to as technologies of governance) and individual self-regulation of behaviour through the internalisation of social norms, which effectively prevent people from seeing that there may be other ways of thinking, talking and acting (Foucault, 1978). For those who seek to "cultivate greater awareness, dialogue, and action toward a post-neoliberal future" (Convertino, Brown, & Wilson, 2015, p. 139), I propose that there are two different ways of responding to the neoliberal fixation with accountability and quality assurance. The first is, of course, simply to disengage from it and focus on building alternative educational institutions with different value foundations, to the extent possible. The second way is to adopt what Nygreen (2017, p. 203) terms a "discourse of both/and", namely preparing students for high-stakes tests (i.e. engaging with the entrenched accountability rhetoric on its own terms) while simultaneously engaging in separate educational undertakings informed by competing values, e.g. education for liberation and social justice rooted in the work of Paulo Freire (1970, 1974). The third way, which may be the most challenging, is to play the game of quality assurance but then attempt to subvert it by changing the definition of 'quality'.

Applying inductive and intersubjective approaches to the assessment and evaluation of education the specific context of the Sustainable Development Goals could generate novel indicator sets and accountability mechanisms, grounded in ethical values and sustainability competencies, rather than relying on inappropriate proxies and reinforcing unsustainable norms. In practice, political will for these efforts may be lacking in the intergovernmental

arena, but there is potential for immediate and far-reaching impact at the grassroots level. Potential arenas of impact include non-formal education initiatives led by civil society (c.f. Nygreen, 2017) (the context in which the ESDinds project originated: ESDinds, 2011); formal 'alternative' schools, such as those belonging to the Steiner Waldorf movement; teacher training colleges in which the sustainability imperative is recognised; and transnational grassroots movements such as the Earth Charter Initiative, which consists of a loose collective of civil society organisations and community groups adhering to a common set of principles outlined in the 2000 Earth Charter (ECI Secretariat, 2010).

By making the success of 'education for sustainable development' (ESD) initiatives conditional on the achievement of 'sustainable development' (SD), a concept which is poorly defined and to all intents and purposes indefinable, intergovernmental institutions are not only embarking on an unwinnable war but also opening up new channels for the consolidation of bureaucratic power. Conversely by defining 'quality ESD' in relation to maximum-diversity sets of stakeholder values, with a particular focus on those with the potential to provide a genuine challenge to entrenched ways of thinking and even to promote revolutionary shifts in consciousness, we can gain a deeper, richer and more nuanced understanding of what `post-neoliberal' versions of ESD might look like and how they can be brought about.

1.4. Methodological Perspective and Methods

1.4.1. Principles of Research through Design

The concept of Research through Design (RtD) was first proposed in 1993 by Christopher Frayling, then Professor of Cultural History and later Rector at the Royal College of Art, in his introduction to the inaugural volume of Royal College of Art research papers (Frayling, 1993). Frayling made a threefold distinction, which he attributed (without citation) to Herbert Read, between 'research *into* art and design', 'research *through* art and design', and 'research *for* art and design'. While Frayling does not actually offer a definition of 'research through art and design', he provides three illustrative examples, namely (i) materials research, such as the colorization of metals; (ii) development work, e.g. customising a piece of technology and documenting the results; and (iii) action research, in which a practical

40

experiment in a studio is systematically documented in a research diary and then contextualised by means of a report.

Over the past two decades, Research through Design has grown in popularity, even to the extent of becoming a 'buzzword' within certain fields of design (van de Weijer, Van Cleempoel, & Heynen, 2014). In architecture, for example, it has been adopted enthusiastically by institutional bodies as well as by leading practitioners (Staff, 2009; The Australian Institute of Architects, 2009; Till, 2008); while in the field of human-computer interaction, it is a recognised method for interaction design (Forlizzi, 2014; Zimmerman et al., 2007). However, the relative lack of interaction between researchers in these two fields and in other domains in which RtD has been utilised, such as industrial design (Pedgley, 2007) and music technology (Barrass, 2016), has hindered the development of a coherent conversation on RtD as a methodology and specifically on what constitutes best practice. Through a broad literature review that cuts across these separate disciplines, I have identified three main questions that can be posed to assess the quality of RtD work, namely:

a)   Is it *research* in the first place?

b)   Is the design process, and the attendant chain of reasoning, *recoverable* from the written account?

c)   Is it *relevant* to a real-world problem, and capable of generating positive change?

The literature relating to each of these questions, which we can term the 'Three R's' of RtD quality, is presented below.

(i)   Is it research?

Frayling (1993) emphasised that documentation of the design process is an essential criterion for classifying a piece of work as RtD: the existence of an artefact does not, in itself, constitute sufficient evidence that research has taken place. Bruce Archer (1995:12-

13), revisiting Frayling's threefold typology of research studies[7], takes this further by setting out specific criteria that must be met in order for a practitioner activity to be defined as 'research', as follows:

> 'One has to ask: Was the activity directed towards the acquisition of knowledge? Was it systematically conducted? Were the findings explicit? Was the record of the activity 'transparent', in the sense that a later investigator could uncover the same information, replicate the procedures adopted, rehearse the argument conducted, and come to the same (or sufficiently similar) conclusions? Were the data employed, and the outcome arrived at, validated in appropriate ways? Were the findings knowledge rather than information? Was the knowledge transmissible to others? Oniy when the answers to all these questions are in the affirmative can a practitioner activity be classed as research.'

Elsewhere in the text, Archer also highlights the necessity of citing primary sources of evidence and, where secondary sources are employed, giving due credit through citation.

In short, as summarised by Nigel Cross (1995), the defining features for classifying an activity as RtD (rather than just a design activity) are that it is *purposive, inquisitive, informed* by earlier relevant research, *methodical* in both planning and implementation, and *communicable* to others in such a way that the results are testable.

In applying these rather conventional criteria (derived from other domains of research) to RtD, it is important not to lose sight of another of its key principles, namely the concept of *embodied knowledge.* This means that the designer's assumptions about the nature of the problem and its potential solution(s) are not only stated openly in writing, but also embodied in the designed artefact (Zimmerman et al., 2007). The research element is thus inherent in the successive iterations of the artefact itself, as well as in the written materials created by the research team to describe the artefact and the process of its

---

[7] A footnote to the electronic version of this paper, included by the transcriber, cites an undated work by Norman, Heath and Pedgley as saying that Archer (an RCA colleague of Frayling) claimed in 1999 to have 'coined' the threefold distinction himself in the 1970s.

creation. In this way, each new version can be seen as a contribution to an evolving conversation: researchers learn through a process of 'dialogue' with the artefact, as it evolves from one version to the next (Toetens, 2013).

Barrass (2016:72) goes as far as to argue that RtD can produce 'richer and more situated understandings' than traditional analytical research, in the sense that 'contributions to knowledge are manifested in artifacts (*sic*) that have much more detail than written or diagrammatic documentation'. This enhanced understanding can be achieved, Barrass explains, by teasing out the similarities and differences between multiple designs in a collection and documenting this comparative process systematically in the form of an annotated portfolio.

(ii)   Is it recoverable?

In contrast to scientific research, it is meaningless to use 'reproducibility' or 'replicability' as a criterion for validity in RtD, for two reasons. Firstly, RtD cannot be conducted on an objective, value-free basis because the investigator is a significant actor in the situation being studied – which, in turn, makes the research situation-specific and impossible to generalise (Archer, 1995). Secondly, as Zimmerman et al. (2007) have noted, it is unreasonable to expect that two designers presented with the same problem will create similar artefacts, even if the problem is framed in the same way each time. Accordingly, drawing on action research literature, it may be more important to talk of *recoverability* than replicability – ensuring that the process can be 'recovered' by anyone who wishes to subject it to critical scrutiny (McNiff, 2013) and that the chain of reasoning leading to the creation of each artefact is strong (M. A. Biggs & Buchler, 2007).

One important aspect of recoverability in RtD is for the researcher to state their 'theoretical, ideological and ethical position' clearly from the outset, in common with other forms of arts and humanities research (Archer, 1995:11). Another concerns the documentation process itself. Pedgley (2007:473), for example, makes several recommendations for documenting a design process to optimise recoverability, namely

ensuring that diary entries are chronological, clear and succinct; that still and moving images of modelling outputs are captured, where applicable; that all diary entries and images are numbered, dated and cross-referenced; and that any 'out of hours' designing is accounted for in the next day's diary entry.

(iii)   Is it relevant?

In addition to meeting the minimum criteria to qualify as research, and being documented in such a way that the design process is recoverable, it has been widely recognised that another essential quality criterion for RtD work is 'relevance' - both in terms of being relevant to a body of theoretical knowledge, and in terms of contributing to the advancement of practice (van de Weijer et al., 2014; van der Hoeven, 2011; Zimmerman et al., 2007).   In this respect, Forlizzi, Zimmerman and colleagues highlight the transformative ideals of RtD – describing it as a process of creating artefacts that are intended to transform society, or to achieve a desired transition from a problematic or undesirable current state to a preferred future state (Forlizzi, 2014; Zimmerman & Forlizzi, 2008; Zimmerman et al., 2007), as well as explaining why the latter state is preferred (Zimmerman et al., 2007, p. 400).

The question of relevance flags up an important concern, which (to the best of my knowledge) has not yet been explored in the literature.   As RtD is fundamentally an interdisciplinary and transdisciplinary activity (Zimmerman et al., 2007), its respective contributions to knowledge and practice need to be problematized, with both 'contributions to knowledges' (in the sense of different bodies of theoretical knowledge) and 'contributions to practices' being understood in the plural rather than the singular.   The respective questions of 'relevance to whom?' and 'relevance for what?' thus become pertinent, especially with the recognition that some voices and discourses may be more privileged than others.   The category 'practices' may refer not only to the evolution of design practice, but also the role of designed artefacts in transforming the day-to-day professional practice(s) of their users, who may or may not be designers themselves. Similarly, a designed artefact does not embody 'design knowledge' alone: it may also constitute a material expression of new knowledge within some or all of the disciplines that have contributed to its creation.   Placing design at the hub of a long-term transdisciplinary collaboration involving both practitioners and

academics, as in the case of the ESDinds project (ESDinds, 2011) and the subsequent 'Evaluating Intangible Legacies' project funded by the UK Arts and Humanities Research Council (Hoover et al., 2015) makes it possible for all three dimensions of RtD quality - academic rigour, recoverability, and transformative relevance or impact – to be pursued simultaneously.

1.4.2. Wider Methodological Perspective

The umbrella of Research through Design encompasses a vast diversity of practices and methods, with the '3 Rs' of rigour, recoverability and relevance providing a helpful set of boundaries for this collaborative exploration space. It is important to note that Research through Design does not invariably involve collaboration: indeed, some of the original studies described by Archer and Frayling appear to have been carried out by single practitioners. Nonetheless, it is at the interface between Archer's original 'research through practitioner action' and what has become widely known and celebrated as 'action research' – as outlined, for example, in the 'Manifesto on Transformation of Knowledge Creation' signed by 60 advisory editors of the journal *Action Research* (Bradbury Huang, 2009) – that this thesis is situated. One of the core principles of action research, as outlined in the Manifesto, is 'partnership and participation'. I have explored the meanings of 'participation' and its assessment in relation to four dimensions – depth, breadth, scope and quality - in two other published papers, supplementary to those included in the thesis (Burford, Kissmann, Rosado-May, Alvarado Dzul, & Harder, 2012; Harder et al., 2013).

Another aspect of the methodological perspective of my work, and beyond that, its ontological and epistemological standpoint, is that of *pragmatism.* In developing this work, I have drawn on the 'pragmatist mixed methods' orientation espoused by Martina Feilzer (2010), which suggests that the most suitable or appropriate methods for research are those that most effectively answer the research questions to generate socially useful knowledge (Onwuegbuzie & Leech, 2005). This avoids a dogmatic adherence to either qualitative or quantitative methods. A pragmatist mixed methods approach situates 'validity' in the theoretical justification for integrating methods, as well as judging both the validity of

qualitative and quantitative methods on their own terms rather than expecting one to conform to the standards of the other (Onwuegbuzie & Johnson, 2006).

A straightforward 'pragmatist mixed methods' orientation acknowledges that there is always more than one possible way of interpreting a situation, but the optimum criterion for choosing one interpretation over another is practical utility. This, however, brings us back to the problematic questions of 'useful for what?' and 'useful to whom?' that can only be answered from a particular - and clearly specified - ethical and cultural standpoint. To address them, co-designers and co-researchers must become conscious of their own biases and prejudgements and make them explicit and transparent. However, this does not preclude them from temporarily setting some of these biases and prejudgments aside– a process known as 'bracketing' (Tufford & Newman, 2010) – in order to achieve a desired goal. In this thesis, for example, I have bracketed the recognition that ethical values are essentially intangible, in order to accept the validity of measurable (or semi-measurable) indicators as proxies for 'values' (see Paper 2). In RtD terms, this is a practical way of creating artefacts that can transform a problematic situation (a 'business as usual' scenario, in which the measurement of global progress and human wellbeing are based on much more inappropriate proxies) into a preferred situation (a scenario in which people are enabled, at least to some extent, to 'measure what they treasure').

In taking this position, I am adopting a methodological perspective rooted not only in pragmatist mixed methods research, but also in *transformative* mixed methods research (D. Mertens, 2010; D. M. Mertens, 2011). As I explain more fully in other published works (Harder, Velasco, et al., 2014; Podger et al., 2013), Patti Lather's (1986) concept of 'catalytic validity' posits that the validity of research is judged ultimately by its success or failure in achieving a desired change. As I have hinted in previous sections, the underlying motivation for my work is grounded in a vision of profound social, environmental, economic, cultural and spiritual transformation. More specifically, the overarching aim of this research is to awaken people's latent abilities to co-design and co-enact strategies for change, with a view to co-creating lifestyles that combine improvements in human well-being with reductions in ecological footprint. In the discussion section, I will examine some of the specific biases and filters that shaped the original ESDinds project; explain how and why these limited the scope and impact of the project; and set out recommendations for an expanded program of

research with a maximum-diversity orientation, which would be more profoundly transformative, transgressive and transdisciplinary ('triple-T') and has the potential for a much greater impact.

1.4.3. Specific Research Questions and Methods

The overarching research question addressed by the four published works in this thesis is how inductive and intersubjective processes, aimed at understanding what is perceived as valuable, meaningful and worthwhile by different stakeholders in relation to a specific project or initiative, might contribute to the development of new strategies for assessment and evaluation in education for sustainability.

Each of the published works addresses this broad question from a different angle, focusing on specific sub-questions and contexts. Research methods have been chosen in accordance with the overarching methodological perspectives described above, with a view to answering these research questions, as set out in Table 1-1 below.

According to the principles of Research through Design, the answers to these research questions are embedded not only in the published works themselves, but in the various iterations of designed artefacts (such as indicator sets, assessment tools and guidelines for their application) that advance transitions from problematic to preferred situations. These artefacts can be found in the final reports for the ESDinds project (ESDinds, 2011) and the 'Starting from Values: Evaluating Intangible Legacies' project (Hoover et al., 2015) respectively, as well as three separate toolkits relating to the use of values-based indicators in Education for Sustainability that were designed, as part of the ERASMUS-funded PERL (Partnerships for Education and Research about Responsible Living) project, for use in secondary schools (Dahl et al., 2014a, 2014b, 2014c).

Table 1-1. Research questions and methods for the four published works

Published work	Research question(s)	Methods used
Chapter 2: Organizational impacts of values-focused evaluation	Which organisational impacts can result from the use of values-based indicators in organizations that regard themselves as offering values-centred approaches to education for sustainable and responsible living, with or without the label of 'ESD'?	Multiple case study method: qualitative thematic analysis of semi-structured narrative interviews, project documents and field notes from trials of values-based indicators in eight organisations
Chapter 3: Operationalising the 'missing pillar' of sustainability	How can the concept of intersubjective conceptualisation, within a clearly defined practical context contribute to the theoretical debates around the 'measurability' or 'immeasurability' of values?	Theoretical analysis of the respective logical arguments for the immeasurability and intersubjective context-specific measurability of values, with a case study of the application of values-based indicators for evaluating an online course for sustainability leaders

Published work	Research question(s)	Methods used
Chapter 4 (Improving indicator design for complex Sustainable Development Goals)	Is there any difference between the approach used to generate indicators for the Sustainable Development Goals, and the values-based approach adopted by the ESDinds project, in terms of the ability of the resulting indicator sets to effectively operationalise the SDG target on 'Education for Sustainable development'	Conceptual framework development through literature review; content analysis of both datasets in relation to the framework
Chapter 5: Promoting sustainability skills in schools	How can understanding the values of educators, across different cultural contexts, contribute to the development of toolkits for helping students in formal secondary education to identify and close 'value-action gaps'?	Multiple case study method: qualitative thematic analysis of semi-structured narrative interviews with UK and Irish conference attendees, UK university lecturers, and Tanzanian secondary school teachers, with a subsequent field trial of candidate indicators and activities in a UK secondary school

Chapter 2:

The organisational impacts of 'values-based' evaluation: preliminary findings from eight organisations

Abstract

Ethical principles constitute a crucial area of debate and discussion in the global conversation around transitions to sustainability, and of particular relevance to the contribution of businesses and other organizations. Scholars in business ethics have recently identified several challenges in this area, such as problems of measurement, rigor, and meaningfulness to practitioners; corporate social responsibility; and institutionalization of ethics in businesses. In this paper, the impacts of a pragmatic values-based evaluation approach originally developed in another field—education for sustainable development—are shown to strongly contribute to many of these challenges. Impacts found across eight organizations include (i) deep values conceptualization; (ii) increased esteem (iii) building capacity for assessment of values-based achievements; (iv) values mainstreaming; and (v) effective external values communications. It seems that the in-situ development and use of values-based indicators helped to conceptualize locally shared values that underpin decisions, thus embedding the application of (local) ethics. Although this study is exploratory, it is clear that the values-based approach shows promise for meeting key challenges in business ethics and wider sustainability, and for new directions for future cross-disciplinary research.

Keywords

Chapter 2. The organisational impacts of 'values-based' evaluation: preliminary findings from eight organisations

2.1. Introduction: Values, ethics and sustainability in organisations

It is acknowledged that the definition, pursuit and assessment of "sustainability" is not only a technical and political issue, but also a moral and ethical one (Clugston, 2011; Hedlund-de Witt, 2011; McCool & Stankey, 2004; Vucetich & Nelson, 2010). Ethical values such as compassion, integrity, justice and respect, and ethics-based decision-making, underpin every aspect of sustainability, including several that are not fully encompassed by the traditional threefold definition comprising environmental, social and economic aspects (Burford, Hoover, et al., 2013; Dahl, 2012). A number of authors propose a "missing pillar" or neglected dimension of sustainability (Burford, Hoover, et al., 2013; Dahl, 2012; Elliott, 2012; Littig & Griessler, 2005). It has variously been characterized as "cultural/aesthetic" (De Leo, 2012; Hawkes, 2001; Kagan, 2011; UNESCO, 2013) or with a focus on Indigenous communities (Nurse, 2006; UN-PFII, 2006; Woodley, 2006); "religious/spiritual" (Clugston, 2011; Hedlund-de Witt, 2011; Interreligious Statement Towards Rio+20, 2012); and "political/institutional" (Pfahl, 2005; Spangenberg, Pfahl, & Deller, 2002). The concept of "ethical values" has been suggested to offer some common ground between these complementary perspectives, and argued as a fourth dimension of sustainability of at least equal importance, inseparable from the others. Difficulties in articulating and measuring the values dimension have been cited as a reason for its neglect, but recent work claims that these can be overcome (Burford, Hoover, et al., 2013)[8].

There have also been a number of high-level calls for the establishment of a global ethical framework for sustainability such as by the Earth Charter (Clugston, 2011; Corcoran, Vilela, & Roerink, 2005; ECI Secretariat, 2010), the United Nations Millennium Declaration (United Nations General Assembly, 2000) and the Earth Systems Science Partnership (Biermann, 2007). More recently, the Club of Rome's "ValuesQuest" program, linked to the United Nations Culture, Creativity and Values Initiative, has explicitly sought to embed ethical

---

[8] This work, published in 2013, is included here as Chapter 3.

values as a key concern in international development discourse (Palmer & Wagner, 2012). Thus, dimensions of values are increasingly being developed in sustainable development.

Organizations and businesses have a central role in the challenge of developing sustainable societies. Already at the turn of the century, Carroll predicted that ethical approaches to business would become a central concern, and emphasized the need for normative approaches to understanding values rather than mere values clarification or "ethical relativism" (Carroll, 2000, pp., p. 41). Over a decade later, mission statements, guiding principles, moral standards and Corporate Social Responsibility (CSR) practices and policies, have become commonplace and core to business activities (Holland & Albrecht, 2013; von Groddeck, 2011) and within them, activities related to values have become increasingly popular (von Groddeck, 2011). Nevertheless, challenges to the application of ethical values in organizations remain. In 2013, Holland and Albrecht surveyed 3600 members of business ethics societies and networks to identify key future challenges for the academic field of business ethics research (Holland & Albrecht, 2013). The results included CSR; perceived challenges with legitimacy and credibility of the field; problems of measurement, rigor, and meaningfulness to practitioners; decline of ethical behaviour; and the institutionalization of ethics in businesses. We return to these issues later in this paper, but here note the overlap of several current issues in wider sustainable development, such as the actual application of ethical behaviours and developing measures for values dimensions.

In this paper, the authors wish to communicate and explore insights from incidental findings from work in one field—sustainable development—which they show to have potentially significant relevance to current gaps in another field of particular relevance to research in sustainability: business ethics. Evaluation interventions were carried out to assist eight organizations to articulate and evaluate their values and related achievements. Anecdotal feedback suggested that impacts of those interventions included a stronger ability to articulate values and embed these in their organizational practices, and so a follow-up study was carried out to explore this. The authors present findings from that study in this paper.

2.1.1. Paradigms in Business Ethics Research

Prior to presenting the study, the authors outline the context of current research in business ethics. A very useful overview of paradigms present in and across business ethics and organization theory has been published by Heugens, Pursey, et al. (2010), charting not only their boundaries and divisions but also the artificial extent of those boundaries, caused by the historical development of various semi-estranged research communities. Although the main intention of that work was to draw out the symbiosis between business ethics and organization theory, a secondary aim was achieved in showing that the pluralism of paradigms and approaches currently present produced a double-edged sword: flexibility and resilience via the range of tools available on the one hand, and a lack of coherence or orientation due to ongoing incompatibilities and even incommensurability on the other hand. The authors illustrated their insights with a multi-faceted presentation of four major themes in the discipline—values, society, power and organizations—from each of modern, symbolic and postmodern viewpoints. Two themes which are of particular interest to us in our own work—values and organizations—were shown to be viewable as individualist or collectivist, almost entirely positivist or constructionist, or anywhere in between, depending on the scholarly community of the researcher. The authors ended with a call to arms for researchers to make greater efforts to engage in science as a social practice and jointly work together across academic "tribes", thus accessing rich sources of new knowledge.

The complex paradigmatic status of the field of business ethics as portrayed by the two studies described above illustrates the difficulties that researchers from other disciplines, such as sustainability, may encounter when trying to communicate concepts and findings which may be of potential interest. This is further complicated by the fact that sustainability research itself is characterized by a diversity of paradigms and perspectives (Pryshlakivsky & Searcy, 2012; Robinson, 2004; Waas, Hugé, Verbruggen, & Wright, 2011). One way of negotiating this inherently messy interface, is to focus initially on pragmatic considerations before expanding in the various theoretical domains (Ioppolo, Cucurachi, Salomone, Saija, & Shi, 2016). Holland and Albrecht—in emphasizing the need for an appropriate blend of academic rigor and practical relevance—lean towards a call for a pragmatist paradigm, stating

that sustainability, the decline of ethical behaviour and globalization are all candidate issues replete with practical questions of ethics in need of answers (Holland & Albrecht, 2013).

2.1.2. Purpose, Paradigm and Context of This Study

The purpose of this work was to carry out a follow-up study to explore incidental findings of values-based evaluation interventions in eight organizations. The findings of the exploratory study potentially pave the way for new directions in business ethics research in particular, and sustainability more generally, through innovations for firmly embedding values (the fourth pillar of sustainability) into organizations and businesses through the application of localized ethics. They indicate a new way of approaching currently identified research gaps, and allow for scoping out and making recommendations towards specifically designed future studies.

The evaluative interventions of interest were designed as part of a European Union FP7-funded research project to facilitate organizations to evaluate values-related dimensions of their work in a manner that they considered valid and relevant. The project used an emancipatory action research approach, working with organizations and businesses engaged in providing non-formal education for sustainable development (ESD). The latter considered themselves unable to articulate "intangible" aspects of their work which they nonetheless considered key, and the work focused on co-developing an approach that could be highly localized (but also generalized and transferable to some extent) (Burford, Hoover, et al., 2013; Burford, Velasco, et al., 2013; Harder, Velasco, et al., 2014; Podger et al., 2013; Ribeiro et al., 2016) which later became known as `WeValue'.

Anecdotal evidence suggested that the WeValue evaluative interventions not only enabled participants to articulate values as individuals and groups, but also led to organizational impacts which far exceeded expectations of researchers or collaborators. Considering the potential importance of effectively bringing values into organizational practices, the authors identified the need for a systematic follow-up study to explore the impacts of the WeValue intervention within a broader organizational context (rather than

focusing specifically on project evaluation in relation to non-formal ESD, which has been discussed in depth by Harder et al. (2014)). It is the findings of this follow-up study, based on a re-analysis of the original dataset, which we report in this paper. The study is necessarily exploratory as the original project was not purposively designed to explore the impacts, but only to investigate the feasibility of developing and using values-based indicators. However, the re-analysis yields important lessons, as well as much material, for the future design of a systematic study.

The study identified impacts seen in eight organizations, from the time of the intervention up until 3–6 months later, and which included greatly clarified understanding and awareness of shared values, the mainstreaming of those values into work, and a new ability to self-develop ways of capturing "measures" of values-related achievements and communicating them more widely. The depth, scale and consistency of the results were noteworthy. Through increasing our understanding that values awareness is an interpretative process where an individual recognizes moral relevance to their situation (Reynolds, 2008), and that moral identity is a precursor to ethical considerations (McFerran, Acquino, & Duffy, 2010) we understood that these observed impacts are very relevant to the raising and embedding of ethics considerations into organizations, and to almost every issue identified as a current challenge in business ethics by Holland and Albrecht (2013). Such a focus on values can provide a useful bridge between individuals, organizational culture, practices and behaviours, and indeed be seen as an integral part of organizational culture which provides unconscious guides for tackling complex ethical issues (von Groddeck, 2011).

This exploratory study purposefully follows the call outlined above for research approaches that focus on pragmatic questions. In doing so, the authors examined the impact of the evaluative interventions through detailed case studies, not assuming specific paradigms or theories. However, it is expected that its results will inform future studies designed to test and link those. In this paper, we will focus on the impacts of the evaluative interventions, but we will begin with a brief description of the design and content of these interventions. The methodology of the exploratory impact study is then described, followed by the findings. We then present a discussion of the findings in relation to three of the named

challenges in current business ethics research—institutionalization of ethics, problems with measurements, rigor and relevance, and CSR—with comments on related other topics.

2.1.3. The WeValue Evaluative Intervention: Developing and Using Values-Based Indicators in Organizations

The WeValue evaluative interventions were carried out in the context of a research project to trial the usefulness and validity of values-based indicators. In this context, it comprised of a number of steps, or variations thereof, delivered in the form of meetings, focus groups, workshops or evaluation activities. The format was not fixed, responding to the different contexts and cultures of each organization, but the core elements are summarized here.

After initial familiarization with the organization and building of rapport, researchers initiated conversations about values in the organization, either with leaders or representative groups. Next, they were asked to look through a selection of values-based proto-indicators previously developed, noting those that resonated with their views on what was important to their own organization. Table 2-1 lists examples of these proto-indicators, which can be variously associated (or not) with value-labels such as Empowerment, Justice, Respect and Care for the Community of Life, Collaboration in Diversity, Justice, Trust/Trustworthiness and Integrity. Any such associations are locally relevant and not found to be transferable to other organizations, but the proto-indicators themselves were found to be useful and transferable as they were designed to be contextualizable for different organizations. For example, the recurring word "people" can be locally defined as any appropriate stakeholder group, e.g., managers, staff or participants/clients, and could be changed to a more specific term at the users' discretion. The word "entity" typically meant the organization as a whole, but could also be used to refer to a specific department, team, working group, etc.

During this stage participants usually found articulations of their initial group thoughts on the list, as well as surprising items which represented values they held but which they were previously not very aware of—causing some "Eureka!" moments. The discussions led to clarifications about proto-indicators on the list (values-in-language) and actions in the workplace (values-in-action). After a considerable period of discussion, participants were

facilitated to prioritize and sometimes cluster the most important proto-indicators to produce their own "list of values-based indicators". This comprised Part 1 of the WeValue approach, which usually had to be separated from Part 2 by at least a few days because the participants often experienced significant clarification of their organizational values, and time was needed for acclimatization: the values-in-action which emerged in the workshops were often newly articulated, or even different to those already formally espoused.

Table 2-1. Examples of 'trigger' values-based proto-indicators used in Part 1 of the WeValue approach

Proto-Indicator Examples Taken from the WeValue "Trigger" List
• Decision-making takes into account the social, economic and environmental needs of future generations
• People are able to suspend their own standpoints during dialogue and listen to those of others
• People feel that their own individual identity and approach is respected
• People feel that they have an equal opportunity to express their opinions
• People are perceived to be transparent
• People are taking the opportunity to develop their own visions and goals for projects and/or for the whole entity
• People investigate what is right and good by themselves, rather than adopting other people's opinions
• Partners trust that each shares a commitment and willingness to collaborate for a similar vision
• People feel that they create something better or greater as a group than on their own
• People feel inspired by the way that leaders live their principles
• The environment/community of life is celebrated

In Part 2 of the WeValue approach, participants were asked which of their own list of values-based proto-indicators they would be interested in developing measures for, depending on their immediate priorities. When identified, the researchers facilitated the participants to co-develop various assessment methods in locally appropriate ways in order to develop "measures" of those proto-indicators that held relevance and validity for the organization. The approach taken was one of co-design and action research, with a strong focus on face-validity. The participants were supported, in effect, to embed ongoing assessments of their own values in their regular activities at levels of rigor appropriate to the purpose required, for example sometimes for self-evaluation and sometimes for external scrutiny. Finally, participants and researchers carried out actual assessments, sometimes few

and sometimes many, and with varied levels of participation in the collection and analysis of data. All cases included a follow-up from the research team.

Table 2-2.[9] A full list of the values-based indicators and corresponding assessment tools developed by one organization, 'DB', during the WeValue evaluative intervention. This organisation started with the two core values of *Respect and Care for the Community of Life*, and *Collaboration in Diversity*, from which they developed 22 indicators, and prioritised those above for development of local measures in their self-evaluation. For more detail on the spatial survey tool, see Harder et al. (2014).

Indicators Developed by One Organization, Which were Chosen for Developing Local "Measures" for Self-Evaluation	Assessment Tools
Partners, member organizations and individuals do not feel that they have compromised their beliefs by participating in the vision and activities of the organization/project.	Spatial Survey
	Focus group
Different points of view are heard and incorporated	Spatial Survey
Degree to which members/partners feel that their individual identity and approach has been respected.	Spatial Survey
People are encouraged to reach their potential	Spatial Survey
Everyone has his/her place in the team.	Spatial Survey
Learning processes accommodate different learning styles	Spatial Survey
Individuals have a feeling of a unified work environment	Spatial Survey
Individuals learn together, share skills, abilities and information freely with one another regardless of creed, color, ethnicity, gender	Spatial Survey
Members are inclusive (talk to everyone and no one is left out)	Spatial Survey
Individuals have a feeling of harmony and pleasant work environment.	Spatial Survey
Long term commitments to protect the environment are created and adhered to	Unobtrusive measures
Education is undertaken to raise awareness and capabilities for the organization to act according to principles of environmental sustainability	Unobtrusive measures
Quality of process and results of activities or projects aiming to achieve or promote environmental sustainability	Forum Theatre Comprehension Test
	Unobtrusive measures
	Key informant unstructured interview
Activities initiated and completed in the conscious aim of contributing to a greater respect for nature	Indirect Measures
Activities initiated and completed in the conscious aim of contributing to a greater understanding and respect of how nature is organized (systems and cycles)	Indirect Measures
Activities initiated and completed in the conscious aim of contributing to a greater valuing of the natural world as a source of personal fulfilment	Indirect Measures
Activities initiated and completed that share with others how to protect and restore the Earth's health	Indirect Measures
The project's activities/events have an emotional effect on participants	Indirect measures
	Word elicitation
	Spatial Survey Indirect measures

---

[9] This table was initially included as Table A1 in Appendix A of the published paper. I have embedded it in the main text here for the purpose of clarity and convenience, and to avoid confusion with other appendices.

As this paper is focused on the impacts of the WeValue evaluation intervention rather than their content, we do not go into further detail here except to mention two notable points. First, the proto-indicator list was derived from in-depth case studies of five organizations (with multiple sub-projects) and then combined inter-subjectively [38]: the character of that list deserves further research.

Secondly, although the proto-indicators were originally derived from certain values labels such as Trust, Integrity and Unity, such affiliations were not usually found useful or necessary in later work. Organizations often made use of the WeValue approach by instead going directly to the list of proto-indicators. In either case the discussions led to a convergence of agreed values-in-action with specific localized indicators.

More information is available on the origin of the research that produced the WeValue approach—originally designed to help civil society organizations evidence achievements that they felt were undervalued (Podger et al., 2010). The action research approach used in its development, which particularly highlighted the importance of localizable proto-indicators, has been reported (Podger et al., 2013), as have several examples of its use in different organizations as an evaluation approach (Burford, Hoover, et al., 2013; Burford, Velasco, et al., 2013; Harder, Velasco, et al., 2014). None of these report on the impacts of the use of the intervention, or insights of any links with business ethics and current related research gaps.

2.2. Methodology

This paper presents a study of the impacts resulting from field trials of WeValue evaluative interventions conducted by university-based researchers in eight diverse organizations. This represents an exploratory re-analysis of an existing data set collected from the participating organizations at the time of the original funded project. It uses a multiple case study approach (Cohen, Manion, & Morrison, 2011) based on a qualitative thematic analysis of the original dataset, comprising semi-structured narrative interview transcripts and project documents. The cases analyzed in this study were selected on the basis of having at least two different data sources available for a particular organization, of which at least

one was a semi-structured interview. Table 2.3 provides an overview of all the data sources used in the analysis.

Table 2-3. Overview of case study organisations and data sources used in the analysis

(all names are pseudonyms)

Organization Pseudonym and Type	Location	Data Sources
"DB": small non-governmental organization promoting environmental education and reforestation through schools and youth groups	Mexico	Two separate face-to-face interviews with Maria, DB director; Formal written report prepared by Carlos, lead ESDinds researcher; Meeting minutes and notes
"FDJ": secretariat of global umbrella organization promoting online sustainability leadership training	Costa Rica	Face-to-face interview with Luisa, project manager responsible for online training; Formal written report prepared by Luisa
"GH": a financial services company without an explicit commitment to ethical values	Luxembourg	Skype verbal interview with Philippe, lead ESDinds researcher; Formal written report prepared by Carlos, assistant ESDinds researcher; Meeting minutes and notes
"HV": a university cross-faculty environmental program	Mexico	Face-to-face interview with Carlos, ESDinds researcher; Formal written report prepared by Carlos; Meeting minutes and notes
"JGSD": global umbrella organization promoting humanitarian aid and values-based youth programs (through a federation of national societies)	Switzerland (secretariat); Jordan (summit)	Skype verbal interview and Skype text interview with Rachel, JMSD project officer in Principles & Values Department (re TMSD project); Formal written report prepared by Antonia, ESDinds researcher (re youth summit)
"MJ": a values-driven company producing and distributing cosmetics	Italy	Face-to-face interview with Stefan, ESDinds researcher; Meeting minutes and notes
"QU": small non-governmental organization training youth to promote conflict resolution in schools through forum theatre	Germany	Face-to-face interview with Josef, QU director; Skype voice call with Ingrid, project manager; Formal written report prepared by Stefan and Karina, ESDinds researchers; Meeting minutes and notes
"TMSD", a member organization of "JGSD" that provides humanitarian activities and projects for conflict-affected youth	Sierra Leone	Skype verbal interview and Skype text interview with Ibrahim, project manager; Face-to-face interviews with Carlos and Stefan, ESDinds researchers; Formal written report prepared by Carlos and Stefan; Meeting minutes and notes

2.2.1. Collection and Preliminary Analysis of the Original Data Set

The data set from the original ESDinds project included (a) researchers' formal project reports and field notes; (b) transcripts of semi-structured narrative interviews with key informants—organization directors, project managers, or both, as practicable—in participating organizations; and (c) transcripts of interviews with the researchers who conducted the field visits and evaluative interventions. All of the interviews lasted 60–90 min and were conducted 3–6 months after the field visits, with a view to clarifying both processes and outcomes of the interventions. The selection of interviewees was made on a convenience sampling basis within those organizations which, at the time of the original study, had already

completed the WeValue evaluative intervention. Standard ethical procedures for prior informed consent and secure data storage were observed. The interviewers aimed to elicit narratives of processes and outcomes during the field visits in participants' own words, without priming for values conceptualization or "benefits", to avoid compromising validity by imposing preconceptions of what the outcomes might be. This dataset had previously been analysed at a relatively superficial level, firstly to meet reporting requirements for project donors (ESDinds, 2011), and secondly in relation to literature on the uses and influence of evaluation.

2.2.2. Thematic Analysis Methodology

A thematic analysis was carried out in order to identify predominant themes relating to the longer-term outcomes of the WeValue evaluative intervention. The full dataset was analysed using "parallel coding", i.e., first coded separately by two authors[10], and then discussed in joint meetings to develop consensus and finalize themes. The researchers used the computer-aided qualitative data analysis software (CAQDAS) program Atlas.Ti. After the themes were finalized, the researchers conducted a content analysis for each of the themes, to identify more precisely the occurrence of related impacts and outcomes.

The thematic analysis was completed using both open and a priori codes; the latter were taken from literature on values, semiotics and inter-subjectivity (cf. (c.f. McFerran et al., 2010)). This was consistent with the research approach adopted, as pragmatists seeking "a good trade-off between theoretical inspiration and openness toward empirical material" in common with Alvesson and Sveningsson (2003, p. 968). The researchers also characterized the findings in terms of their impact relating to managers, staff, or clients involved in the evaluative intervention (c.f. Agle & Caldwell, 1999). The authors opted not to fully quantify the data, due to concerns that listing the number of case studies in which each outcome was observed might mislead readers into inappropriately ranking outcomes in order of importance on the basis of frequency alone (see Table 2.4).

---

[10] I created the *a priori* codebook and was one of the two coders; the other was Lee Stapleton.

2.3. Findings

Five main impact themes were developed from the analysis, namely (i) values conceptualization; (ii) esteem raising (raised awareness and value of identity at individual and group levels); (iii) values mainstreaming/ internal transformation impacts; (iv) values-based assessment capacity-building; and (v) external communications impacts.

A summary of their relative occurrence at different levels (managers, staff and clients) is shown in Table 2-4, and illustrative quotations are given for each in the sections below and in Table 2-5.

Table 2.4. Themes relating to the impact of the WeValue evaluative intervention, and an overview of the strength of evidence for each. A single tick denotes outcomes observed in only one organization; a double tick in two or more. A question mark indicates outcomes that the authors regarded as likely, but were not demonstrable from the data.

Themes	Managers	Staff	Beneficiaries/Clients
Values Conceptualization			
Adding referents to value-labels	√√	√√	√
Adding value-labels to referents	√√	√√	?
General understanding and acceptance of values	√√	√√	√√
Esteem-Related Outcomes			
Increased self-esteem/self-awareness	√	?	√√
Increased understanding and acceptance of others	√√	√√	?
Increased understanding of the organization as a whole	√√	√√	√
Increased understanding of wider human-environment system(s)	?	√	?
Assessment Capacity Building			
Increased understanding and acceptance of indicators	√√	√√	√√
Increased understanding and acceptance of assessment methods	√√	√√	√
Internal Transformation/Values Mainstreaming			
Personal commitment ("buy-in") to organization's activities	√√	√√	√√
Individual behavior and/or group dynamics	√	√	√√
Values mainstreaming within the arena of:			
Strategic planning processes	√√	√	√
Training and internal communications	√√	√	√
Individual performance assessment	√√	√	√
Organizational performance assessment	√√	√	√
External Communications			
To donor organizations	√	?	√
To existing partners or clients	√√	√	?
To prospective partners or clients	√	?	√
To policy-makers	√	?	?

Table 2-5. A summary of illustrative quotations for each theme

Theme	Sub-Themes	Illustrative Quotations
Values conceptualization		"… all participants stated that their consciousness of the presence and importance of values had been greatly heightened, and that after the field visit they tend to look in terms of values at their work and interactions, both individually and organizationally, in a new way." [44]
		"…in a region full of sexism, where women do not have that access (to information and decision-making) generally, the youth realized that the project has generated a space of equity. However, that (the WeValue approach evaluation) was the moment when they became aware of it … With the youth, I had been working consciously, very much, around providing that equity, but I never gave them a logo about it. I never said "this is about equity", I just created it." [45]
		"I think [the youth] got a better understanding of what it is to be involved in (the JGSD global initiative) and that values are the foundations of this movement … It gives them a global picture and a deeper understanding of how you can put into action those values." [46]
		"Only four JGSD Fundamental Principles were cited by the trainers: humanity (11 times), unity (3 times), independence (1 time) and service (1 time)—compared to respect, which was cited 43 times. This suggests that identifying values is very subjective, but also that the process of identifying values from indicators (more tangible expressions of those values), draws out values that are based on experience rather than ones based on more abstract concepts." [47]
Esteem-related outcomes	Increased understanding and acceptance of self, others, organization, and ecosystem as a whole (all inter-related)	"With respect to the youth, they said it in public, that after this process they understand one another better and they value much more what they're doing. They've always felt very united, but now they know why they're united." [48]
		"Through the processes and assessment tools, we (managers) were able to get a deeper insight into the young people's sense of self, each other and the community. The biggest help for me was an added insight into their motivation, awareness and consciousness of themselves and others, and the connection with the group and the environment." [45]
		"Professor S. stated that she had observed positive attitudinal changes as a result of the field visit in all participating staff and volunteers and in her own self, in a way she felt was deeply rooted and transformational." [44]
		"It was really, really encouraging. (The youth) can see the things that they still have to work on. In addition, then in some places, we could say to them, "There is this gap between you and us (i.e., managers); what can we do about it?" It felt really good. I was the one who led the conversation. It was positive, I really enjoyed it. They became more self-confident." [49]
		"I would say that this has affected them positively in terms of respect for leadership, now they realize that those leaders were not imposing something or showing that they wouldn't improve, but they were the spokesmen or spokeswomen much more." [50]
Assessment capacity building	Increased understanding and acceptance of indicators and assessment tools (inter-related)	"(At a national youth camp, the youth group members) ran the same process for other youth. They offered it as an ice-breaker activity. They did a spiral of "where are you?"(i.e., spatial survey using a spiral)—they chose four indicator questions and asked them, using the spiral." [45]
		"They simply liked the assessment tool, and they want to be able to replicate it on their own, so they taped it on a video … " [51]
		" … what's really good for us is this observation tool. It's simple, easy, objective, this is something we really can use." [49]

Table 4. Cont.

Theme	Sub-Themes	Illustrative Quotations
	Personal commitment (buy-in)	"It raised awareness of where they should be, what they should be doing in their communities as young people who are actually changing their mindsets, especially with these Principles and Values of JGSD…. making a difference within their own communities with their own actions. They seem more mature." [46]
	Individual behavior and group dynamics	"The original indicator was 'Group norms exist and they are followed.' For that question everybody went into (the part of the spiral that represented the answer) 'More or less'. We asked them, 'Why? You make your own rules, and there is no pressure or imposition of these norms, why don't you respect them?' And as a result of that process the youth made a commitment (to respect them), without me putting any pressure on them, but because they felt ashamed. They promised to follow the norms like arriving on time and keeping the blog updated. It has worked, I am here and they are doing everything! That result transformed the group relationship." [45]
	Strategic planning	"As a result of this process we decided that next year we would not do as many activities but we would identify those with the highest impact on selves, communities and ecosystem, thanks to the information provided by the indicators. It helped us to prioritize our activities, we had such a broad spread of action." [45]
Internal transformation; values mainstreaming	Training	"One of the peer educators said that previously, in presenting the Earth Charter in the workshops she used to focus on concrete behaviors, such as recycling waste. Following the field visit she stated that she now puts a much greater emphasis on the Earth Charter values, and sees the Earth Charter not just as a way of achieving specific behaviors but in terms of the development of the whole individual, beginning with herself, and for participants also…. [44] "After (the field visit) we changed our preparation phase (for new volunteers) to values. The first week we focused on service, the second week on consultation, the third on being an example, the fourth on consultation, and the fifth on unity … . Before when we did the preparation phase, we just focused on themes, like acting, etc., and each week we would look at a topic connected to [the organization] and the show." [49]
	Internal communication	"We could see where there were differences between the youth and the staff, especially about information, we could see where the youth especially see that there are difficulties with communication between them and the staff." [49]
		"I realized that before I was categorizing (schools) according to more superficial aspects, and what was missing was the values. That made me realize that we have to impart the value of Respect and Care for the Community of Life right at the start… if that value wasn't there then the commitment or the energy wouldn't be there, even if everything else seemed perfect…Everything has acquired a clearer lens since the visit." [45]
	Individual performance assessment	"(Our previous approach to individual performance assessment) was just my own intuition as a teacher, but without any certitude…I could see values there in the relationships and the commitment, but I couldn't see how it was possible to measure. (In the WeValue approach process) through dialogue, we were gradually arriving at the complexities and translating it into something marvelously simple that gave deep information." [45]
	Organizational performance assessment	"The process helped us to understand how we were doing in relation to assimilation of the values that we promoted as a core goal, both for the entire project and in each of the participating schools." [45]

Theme	Sub-Themes	Illustrative Quotations
	Donors	"Thanks to this process, (our major donor) has re-conceptualized the work of DB as something of international relevance, no longer just a local project." [45] "(DB's major donor) was very interested. They asked me if I would be interested in developing an entire indicator system for their national work, tailor-made, as opposed to the generic WeValue one." [52]
	Existing partners/clients	"In the schools it worked, the process was very helpful, in particular helping the teachers and the headmasters to understand that my work is not just teaching the children how to plant trees, but also in values. It helped in the relationships with the institutions. The schools saw that we were creating respect for the community of life." [45]
	Prospective partners/clients	"It would make it easier for new schools who think about working with us. It's sometimes hard for us to explain what goes on in our performances. We normally have to give them a live example, show them a full performance so that they can see an example of what we do, and then they decide. It works, but it's very time-consuming. If we can measure values, we can give them more clarity." [53]
External dissemination	Policy-makers	"We nearly had a couple of meetings with the President. We did have a meeting with the Minister for Education and Youth who had a behavior change program which the President is very close to. They were very interested in the WeValue approach. That was also part of what helped the process of assimilation. Because at that meeting it wasn't just us presenting it, but also (local managers), particularly 'Ibrahim', presenting the WeValue approach as he understood it to the government, in terms of the relevance that he felt." [54]
		"… since we met our national secretariat here and we spoke about these values indicators and the measurement of these values indicators, and (in) their own jargon now, every time they come on to the media that's all that they're talking about." [55]

2.3.1. Values Conceptualization

Very strong evidence of values conceptualization was found in all eight organizations studied, (i.e., creating or enhancing experiential understandings of "values"). This occurred through strengthening of the links between values-in-language (such as named values-labels like Trust, or specific values-based indicators such as those in Table 2-1) and values-in-action. This occurred in three slightly different mechanisms: (a) starting with the value-labels and then adding "referents"—statements about how those values would be enacted in practice within the given context; (b) starting with the "trigger" proto-indicator(s) and then identifying relevant "values-label" words or related phrases; and (c) by starting with proto-indicators in language and relating them to values-in-action without use of any values-labels in the process.

Whichever values items were used, by reflecting on them individually and then discussing, exploring, debating and modifying the details of them collectively, participants achieved a deeper and richer shared understanding of how these espoused values were (or could be) enacted within their organizations. The data revealed that they took this new perspective with them in the following months:

> " . . . all participants stated that their consciousness of the presence and importance of values had been greatly heightened, and that after the field visit they tend to look in terms of values at their work and interactions, both individually and organizationally, in a new way." (ESDinds, 2010b)

In some cases the values-based indicators which were discussed and refined were similar to the values espoused in pre-existing documents such as mission statements and websites, but in other cases they were very different interpretations or even entirely new "core values" that had not previously been acknowledged or discussed. Several examples were seen of underlying values which were key to group work—but previously unarticulated—being elicited and then verbalized:

> "...in a region full of sexism, where women do not have that access (to information and decision-making) generally, the youth realized that the project has generated a space of equity. But that (the WeValue evaluative intervention) was the moment when they became aware of it . . . With the youth, I had been working consciously, very much, around providing that equity, but I never gave them a logo about it: I never said 'this is about equity', I just created it." (ESDinds, 2010d)

Beyond the distinct processes discussed above, namely the addition of referents to specific value-labels and vice-versa, we also found evidence for a broader, slower and "fuzzier" values conceptualization and awareness-raising effect arising from theWeValue intervention. This seemed to begin with an enhanced general awareness that values did indeed underlie day-to-day actions within the organization, moving to a realization that they could be made visible through values-based evaluation, and thus to a reinforcement that values provided a useful and relevant framework with which to view their work:

> "I think (the youth) got a better understanding of what it is to be involved in (the JGSD global initiative) and that values are the foundations of this movement . . . It gives them a global picture and a deeper understanding of how you can put into action those values." (ESDinds, 2010g)

Although assessment capacity is mentioned below as a separate impact category, it is worth noting the finding that when the organizations performed assessments of their values-based indicators and then had before them "measures" indicating the extent of their presence, that this had a visible reinforcement effect in some cases. The quote about "equity" above is an example: assessment results reinforced that it was present and reminded participants of its importance to them. Thus, assessment reinforced values conceptualization and awareness-raising.

3.2. Esteem-Related Outcomes

The WeValue evaluative interventions contributed to deepening people's understanding,

acceptance and valuing of themselves, one another, and the group or organization. These ranged from improvements in the self-esteem of individual managers or employees, through changes in the way in which managers understand and value their staff (or vice-versa), to new understandings of the significance of the organization's work (see Tables 2-4 and 2-5).

The distinction between general values conceptualization and these esteem-related outcomes may not always be entirely clear. However, we see the former more as the development of a values-based lens with many components, whereas the latter are consequences, e.g., new ways of viewing colleagues and work through that new lens:

"This project helped us a lot and it still helps. We try to think about what is the value behind (participants' actions) . . . This is how we see each human being, full of values." (ESDinds, 2010f)

Some interviewees also commented explicitly, and others implied, that these outcomes have a strong morale-boosting effect. Thus, reflection on one positive outcome could generate others. In one case an organization (DB) was inspired by Indigenous environmental values and the director alleged that the intervention had changed not only the participants' view of their own organization but also their understanding and awareness of the human-environment system within their forest (ESDinds, 2010d). Our qualitative data contain several examples of improved relationships between managers and staff, improved self-esteem, and specifically "feeling understood and valued".

2.3.3. Assessment Capacity Building

In the second part of the intervention (Part 2), participants were facilitated to understand and develop assessment tools to produce "measures" of the indicators they had prioritized. In many cases, assessment tools were designed to fit in closely with the day-to-day activities of the organizations. The degree of rigor and type of validity required was discussed and varied depending on the specific needs. Table 2-2 shows those used for one organization, including observations, surveys, key informant interviews and theatrical performance. Some indicators had only single, informal measures for one aspect, while others

had several assessments, designed to provide a more rigorous evaluation. It was not expected, from the initial intervention, that the organizations would make a sustained use of the assessment methods developed, or be able to go further and develop new ones in the future.

Our findings showed that, several months later, a number of organizations had continued to use—and, in a few cases, even to extend—the assessment methods. It seems that the WeValue evaluative intervention can be useful for building capacity in performance assessment. Managers and their staff (and, potentially, clients) can learn new ways of understanding and evaluating the work of individuals and of the whole organization: in ways which resonate with them and instil a feeling of ownership. Some of the organizations' managers, staff and even clients acquired new understandings of the concept and uses of indicators, and of specific assessment methods (See Table 2-5 for illustrations).

In two cases, DB and TMSD, managers and clients were empowered to use the WeValue evaluative approach independently in different contexts (ESDinds, 2010d, 2010e). As these were non-profit groups that were not previously comfortable with performance assessment, beyond their formal donor-driven evaluations, this was a significant and unexpected outcome.

The experience of the WeValue evaluative intervention inspired organizations accustomed to quantitative evaluation to explore qualitative and creative ways of understanding impact. Conversely, small organizations who have avoided formal evaluation found the WeValue evaluative approach an attractive alternative which encouraged them to adopt systematic surveys and observations. As previously mentioned, assessment of tangible "measures" of values concepts in the organization inevitably reinforced their meaning to participants, embedding the shared values.

Although not seen in these eight examples, this also opens the possibility of reinforcing dissonance in groups where the "shared values" are not shared by all participants.

2.3.4. Values Mainstreaming and Internal Transformation

Considering the existence of the findings discussed above of values conceptualization and awareness raising, increased value of self and others, and increased ability to determine measures of values in everyday activities, it is perhaps not surprising that the authors also found evidence of the joint impact of those into a higher-level impact of values mainstreaming, i.e., the shifting of the organization to a more "values-based management" approach. Evidence was found of observable changes in the following arenas: assessment of individual and/or organizational performance; strategic planning; internal communications; and training protocols.

Internal transformation, while encompassing all of the above, is a broader category that can also include changes in "buy-in" or commitment to the organization and its activities, as well as changes in individual behaviour and group dynamics—not necessarily directly values-based. Increased buy-in, changes to strategic planning processes, improvements in individual behaviour and group dynamics, and changes to training, assessment and internal communications protocols were all observed at the managerial level in more than one case study. At the staff level, increased buy-in was observed in several cases, and each of the other named outcomes was evident in one case study. Among participants/clients (e.g., beneficiary youth groups), both "increased buy-in" and "changes in individual behaviour and group dynamics" were observed in more than one case study, and each of the other outcomes in one case study. Specific examples of all these impacts are given in Table 2-5.

It is possible that by illuminating "value-action gaps", the WeValue approach may serve as a catalyst for self-directed individual behaviour change without the need for authoritarian managerial intervention, a theme that the authors will explore further in future work (and for which we welcome collaboration). Of particular interest is the example of concrete behaviour change cited in organization DB, where the members of the youth group recognized through the participatory evaluation that they were failing to meet their own self-imposed norms. As a result, and without any direct intervention from their manager, they changed their behaviour in order to conform more fully to the standards that they had set themselves. This, in turn, transformed the group dynamics (ESDinds, 2010d).

2.3.5. External Communications

As might be expected from transformational work on organizational values, important subsequent external communication activities took place between organization managers and their counterparts in partner organizations, such as the head teachers of schools participating in an environmental project.

There were also significant and intriguing single-case examples of new styles of communication to donors and policy-makers in organizations DB and TMSD respectively (ESDinds, 2010c, 2010d). Table 2-5 provides illustrative examples from the data. The interview data hint at secondary outcomes occurring within these recipient organizations and institutions, which in turn have gone on to adopt value-labels and the WeValue approach referents into their own vocabularies. The major donor to DB has since commissioned a large multi-level evaluation of its own national programs using the WeValue approach (Podger et al., 2016).

The data shows that new understandings generated through the WeValue approach can catalyse change within the organization in question, and also transform the way in which the organization presents itself to others.

2.4. Discussion

The findings detailed above clearly illustrate how the WeValue evaluative interventions helped organizations respond to a number of key issues identified in business ethics research. Some of the challenges mentioned by Holland and Albrecht (2013) are directly addressed, such as problems of measurement, rigor, and meaningfulness to practitioners; the institutionalization of ethics in businesses; and CSR. We first discuss the linkages to these named challenges, and then consider linkages to some conversations in the wider business ethics literature.

2.4.1. Problems of Measurement, Rigor, and Meaningfulness to Practitioners

Fundamental to the usefulness of the WeValue approach is its ability to efficiently facilitate participants to articulate in discourse what they already undertake or experience together as values-in-action. In that way, they jointly develop values-based indicators—short sentences that clarify a manifestation of a shared value. This process helps to concretize values that underpin decisions about what is or is not ethical. Once produced, the same indicators effectively provided criteria for the practice and application of ethics—which can also be used to devise informal or formal "measures" that are valid to the participants and adapted to the context (see Appendix A). Organizations that had previously eschewed evaluation methods of any kind, enthusiastically developed localized assessment methods to provide "measures" of these indicators which then informed observable practices of individuals and groups—precisely because they considered the indicators very meaningful to them and the measures a mechanism for self-evaluation and learning. The levels of rigor used depended on the purpose: in some cases three assessment methods were used for one sole indicator, which itself might be only one of a set of 10–12 indicators designated—locally, by that organization—to be elements of one core value. The measures developed for those twelve indicators could then provide a meaningful and rigorous representation of that core value to members of the organization, their stakeholders or funders. In other cases, participants used a simple measure of fewer indicators, giving more importance to the localized statement itself rather than an overarching label. Regardless, the purposefully designed localizability of indicators, and flexibility in choice and design of assessment methods, provided local solutions to these challenges for the practitioners involved in these eight organizations. This process does not exclude external evaluations: external evaluators could be allowed to specify other indicators from the locally derived pool, or still use their own.

3.4.2. Institutionalization of Ethics

The institutionalization of ethics in organizations as described by experts in the field of business ethics broadly refers to developing strong ethical cultures and a clear focus on how to apply ethics in all practices of the organization. The findings from this exploratory

study suggest that clearly conceptualizing shared organizational values—through meaningful discussion—and developing concrete indicators to effectively define them in practice, led to mainstreaming of values in the organization. The fact that the indicators were designed to be operationalizable, and that the participants co-developed specific assessment tools for the values-based indicators may be key: they might have effectively integrated different ways of understanding ethical values in everyday practices. The contextually developed values-based indicators became reference items, not only for possible evaluation, but also for guiding ethical judgments during decision-making processes. Furthermore, the participatory nature of the WeValue evaluative intervention, which included deep discussions leading to the conceptualization of shared values, and involvement in the evaluative process, gave space for reflection on developments in the organizations' ethical culture. An example of this was the youth group in organization DB which realized it was not conforming to its own agreed group norm of punctuality, and changed their behaviour—without managerial intervention. The director of the same organization reported presenting their work differently to funders after embedding the new indicators in their work: no longer as an organization focused on reforestation alone, but on reinforcing values such as empowerment, equity, and the emotional connection to the wider community of life.

2.4.3. Corporate Social Responsibility

The values conceptualization, mainstreaming and external communication impacts of the WeValue approach outlined above have important implications for the understanding and application of CSR in organizations. The indicators approach allowed for the conceptualization and operationalization of the organization's values, in turn providing clear guidance for managers, staff and beneficiaries on the actual practices that support CSR within the organization and helping to move "from rhetoric on business ethics and CSR to meaningful action" (Holland & Albrecht, 2013, p. 783). In addition, the values-based indicators and their measures provide potential for generating measures of ethics and CSR practices, and can provide innovative and meaningful ways of communicating such practices.

Values conceptualization and mainstreaming can generate diverse examples of a company's commitment to values—and provide conceptual and strategic frameworks for

reporting CSR. This is a valuable tool for companies faced with an increasingly cynical public wanting evidence of the triple bottom line (Christofi, Christofi, & Sisaye, 2012) and critical of empty values statements not backed up by action (Du, Bhattacharya, & Sen, 2010). Indeed, while 64% of the 250 largest multinational companies published Corporate Social Responsibility (CSR) reports in 2005, the majority of them listed multiple uncoordinated initiatives rather than elucidating a coherent strategy (Foote, Gaffney, & Evans, 2010). The WeValue approach thus appears capable of providing a values-based strategy, and clear indicators for internal and external communication, monitoring and evaluation.

2.4.4. Linkages to Wider Conversations on Values in Organizations

Although a relatively new focus in sustainability research, the concept of values has long been central for conversations business ethics, management and organization studies. We show here some linkages of the impacts of the WeValue approach to some of them.

In an effort to study the challenge of applying ethics in practice, a number of researchers have focused on behavioural ethics approaches, investigating the influence of values held by individuals within an organization, with evidence that these can influence employee behaviour and decision-making processes. Values have been described as a key element of exemplary leadership (Kemaghan, 2003), a way for managers to influence individual behaviour without resorting to authoritarianism (Buchko, 2007), and an important mediator in decisions about equal resource allocation (Garcia, Bazerman, Kopelman, Tor, & Miller, 2010). Enhanced values focus in management has also been shown to increase trust and understanding between managers and their subordinates (Artto, Kulvik, Poskela, & Turkulainen, 2011; Christensen & Lægreid, 2011).

This is in contrast to approaches in organization studies, which focus on values at the organizational level. Such research provides evidence for the influence of collective values— and how these are perceived—on those working within organizations. In their study of 902 managers from different contexts, Huhtala et al. (2011) demonstrate that there is a positive relationship between managers' perceptions of ethical organizational culture and occupational well-being. Other studies have demonstrated the positive effects of congruence

between individual and organizational values on staff retention, satisfaction and increased ethical behaviour (Andrews, Baker, & Hunt, 2011; Verquer, Beehr, & Wagner, 2003). Indeed, values can be seen as an integral part of organizational culture that provide unconscious guides for tackling complex ethical issues (von Groddeck, 2011).

The WeValue approach provides some insights to these studies, and links between them. For example, the impact findings show how collective conceptualization and later evaluation of values-based achievements has effectively united individuals, defining and reinforcing shared values, and collectively clarifying which activities are within or not within their boundaries. That effect could be seen as the focusing of the workforce on clarified mission statements; managers and staff alike. The importance of the elicitation of unarticulated values-in-action into tangible values statements demonstrated in the WeValue approach also suggests that a distinction might need to be made in research generally between those and the more superficial values words used in everyday language without prior reflection. In fact, findings from previous studies might be negated in cases where individuals do not have shared values-in-action to build on but are trying to reconcile superficial values with no grounding in a context-specific reality. In such cases, values-based approaches might not only be non-useful but could easily lead to misunderstandings and even polarization of differences.

The success of the WeValue approach at triggering conceptualization and the production of indicators that are deemed valid by participants suggests it deserves further study, because a key feature of organizational values discourse is its portrayal of values as highly abstract. Values tend to be poorly conceptualized even at the individual level and are often below the level of full conscious awareness (Agle & Caldwell, 1999; Meglino & Ravlin, 1998). In the absence of collectively agreed definitions for specific value-labels such as "trust", "justice" or "integrity", individuals rely on their own tacit or explicit understandings of meaning. These understandings vary according to ethnicity, previous life experiences and current circumstances, and are highly context-dependent (D. Brown & Crace, 1996; Peng et al., 1997). However, the concept of values is arguably useful within the context of business ethics as it provides a bridge between individuals, organizational culture and subsequent actions and behaviours (von Groddeck, 2011, p. 72).

Concerns about the abstract nature of values communication, and its perceived irrelevance to daily life, are raised elsewhere in the business management literature. Gruys et al. (2008, p. 833) lament that "too often the values of organizations show up on laminated cards or wall plaques, rarely heard or seen", while Lencioni's damning verdict on corporate values statements is that the majority are "bland, toothless or just plain dishonest", with resulting destructive impacts on employee morale, client satisfaction and managerial credibility (Lencioni, 2002, p. 113). Values transmitted implicitly through other channels such as rituals, drama, stories and symbolic constructions within an organization's culture may contradict those that are explicitly communicated in oral or written forms (Cha & Edmondson, 2006). Even when there is a prominent and unambiguous values discourse within an organization, shared understanding of value meanings may be lacking. Focusing on values can thus backfire, as Cha and Edmondson (2006, p. 71) explain: "The inherently abstract or "fuzzy" nature of values creates the potential for multiple plausible interpretations of the values' appropriate meanings." They describe the problem of "values expansion": the tendency of employees to add new layers of meaning to the value-labels passed on to them by senior management. This may result in attributions of hypocrisy to leaders even as they continue to remain faithful to their original understandings of the value-labels, and often generates disenchantment—a "toxic" blend of frustration, anger, disappointment and loss of trust— among employees. Cha and Edmondson speculate that such disenchantment might trigger increased absenteeism, impaired job performance, and detrimental effects on work attitudes and behaviours. Similarly, Lencioni (2002) provides an actual example of a company in which incongruence between leaders' and employees' understandings of a single value-label led directly to the resignation of a senior executive.

Another important insight from the WeValue approach is contextualization. Early on in its development, it was clear that it needed to be steered away from being a rigid framework to one which could be localized and thus "owned" rather than seen as an external imposition. This successful localization of values-based indicators is only possible with the contribution provided by members of the organizations when they collectively reflect on their own actions – concretely contextualized in the workplace.

Business ethics concerns itself with the application of certain values that are related to deliberate decisions about what is or is not ethical in a given context. Thus, ethics and related organizational values are arguably more explicit and purposefully linked to practices and behaviours. In practice, this has been translated by the development of codes of ethics as mechanisms for ensuring such practices. However, codes and compliance-based approaches have been criticized and, akin to values discourses, remain vague and lack specificity (Warren, Gaspar, & Laufer, 2014). Furthermore, what is ethical in a given organization is defined contextually and will not depend solely on a code (Christensen & Lægreid, 2011). In this context, there has been increasing interest in understanding ethics as practice, in other words that organizational ethics are constructed through a multiplicity or behaviours, decisions, inter-personal interactions "not only embed and enact ethics but also form the framework for their institutionalization, politicization and contestation" (Clegg, Kornberger, & Rhodes, 2007, p. 94).

This way of understanding ethics may contribute to challenges faced by scholars looking at organizational values. Indeed, although values-based approaches are more effective, they do not necessarily generate tangible long-term outcomes (Warren et al., 2014). The focus merely on the level of values discourse is limiting: related practices can also influence ethical organizational culture and specific behaviours such as willingness to report unethical behaviours.

The relatively limited evidence provided in this dataset suggests, provisionally, that the WeValue approach may contribute to the development of ethical cultures in organizations. Existing literature hints at possible ways to do this: Gruys et al. (2008) suggest, for example, that a potential solution might be to obtain a measure of organizational values enactment by linking statements of espoused values directly to formal performance assessment systems. That measurement-based approach contrasts with Cha and Edmondson's more constructionist recommendation of "thoughtful dialogue" between managers and employees about the meanings and practical implications of value-labels (Cha & Edmondson, 2006, p. 75). In a recent paper, Warren et al. (2014) argue that ethics training that examine practices related to ethical dilemmas in detail have more lasting effects that merely focusing on vague codes of ethics and organizational values. All of these studies are

consistent with the impact findings reported here, which seem to go further and produce an embryonic framework to understand them.

2.5. Conclusions

The successful application of ethical values in organizations is of crucial importance for contributions of business and civil society to sustainable development. Business ethics scholars and practitioners have been at the forefront of such work in the past decades, and have recently identified outstanding challenges and gaps in research. These include problems of measurement, rigor, and meaningfulness to practitioners; CSR; decline of ethical behaviour; and the institutionalization of ethics in businesses. Although this study was designed retrospectively as an exploration of unexpected outcomes of a research project in another area (sustainability indicators), the findings leave little doubt that the approach used in the WeValue evaluative interventions produced impacts that contribute significantly to current challenges in business ethics, starting with increased awareness and the institutionalization of values in the organizations, with related domino effects on ethics in discourse, communication and action.

The analysis provides clear indications of important areas of impact that are likely to be generalizable to other organizations using the WeValue approach, although dedicated designed studies should be carried out to confirm them. First, the approach may help participants to gain a shared understanding of "core shared values" in a more tangible way by collectively associating reasonably specific word-based phrases (proto-indicators), with local meaning obtained through consideration of concrete examples (values-in-action). This greatly raises awareness of local shared values in the context of the organization, which is a requirement to considerations of ethics. Second, the approach may enhance esteem and strengthen relationships at several levels, increasing understanding and acceptance of oneself, other individuals, the organization as a whole, and even wider human-environment systems. This raises the ethical consideration of the human aspects of co-working beyond work roles, and can be seen as the beginning of the embedding of ethics into relationships (with a knock on effect of reducing distance between managers and staff). Third, it may increase understanding and acceptance for various assessment methods, which become

80

considered valid locally, thus building internal capacity for values-based performance assessment (which may be especially important in small non-profit organizations) and an expanding awareness of what is "measurable". Related to this is the reinforcement of the usefulness of values-based perspectives when measures of them are made tangible. Fourth, the use of the WeValue approach may catalyze internal transformation within the organization, both in terms of "values mainstreaming" (i.e., prioritizing specific ethical values in strategic planning, decision-making and performance assessment) and in other respects such as improving group dynamics and promoting positive behaviour change without the need for authoritarian management. Finally, the approach can provide organizations with a new shared vocabulary for communicating previously intangible values-related goals and impacts to key external stakeholders, notably donors, policy-makers and prospective clients, and our data shows that the bridging of this significant communication gap can have immediate impacts of allowing stakeholders to work together more effectively towards shared understandings of goals.

Several caveats and limitations need stating. First, "the" WeValue evaluation intervention as used here was under development and carried out in the context of an ongoing research project, thus allowing for significant time and researcher involvement in the selection of indicators, development of assessments and follow-up. This also meant that elements within the intervention varied between the organizational contexts—both in materials used and facilitation approaches. The WeValue approach has since been further developed, and although the research team believe similar types of impacts would be seen regardless, they have no feel for the variations in strength of impacts that might change. Secondly, although the organizations differed in size, geographic location and nature (civil society/private sector), all were involved because of their connection, direct or indirect, to members of the project consortium, and thus were likely to be interested in values-related work. In future studies, organizations should be chosen more appropriately for an impact research question. Thirdly, the impacts were not predicted or planned for and thus no baselines, measures, or monitoring over time were carried out, and other types of impacts may have been missed completely. Finally, the study does not adequately account for long-term impacts, as the data used in the study was collected a maximum of 6 months after the intervention. All of these limitations need to be considered in any future studies. It is

suggested that the early version of the WeValue approach is replaced by further developed and systematized ones used in a wider diversity of organizations, using values-based indicators from a common reference list deemed suitable for the organizational context. (This can be found via an open search on the internet). Future studies should include baselines and ongoing, long-term, non-participatory measures of the types of impacts reported here, as well as wider aspects. It is also suggested that researchers in business ethics and organizational studies be directly involved in such future work, to allow appropriate expert joint contributions.

Chapter 3:

Operationalising the 'missing pillar' of sustainability: towards intersubjective values-based indicators

Abstract

This paper argues that the need for a core 'fourth pillar' of sustainability/sustainable development, as demanded in multiple arenas, can no longer be ignored on the grounds of intangibility. Different approaches to this vital but missing pillar (cultural-aesthetic, religious-spiritual, and political-institutional) find common ground in the area of ethical values. While values and aspects based on them are widely assumed to be intangible and immeasurable, we illustrate that it is possible to operationalize them in terms of measurable indicators when they are intersubjectively conceptualized within clearly defined practical contexts. The processes require contextual localization of items, which can nonetheless fit into a generalizable framework. This allows useful measurements to be made, and removes barriers to studying, tracking, comparing, evaluating and correlating values-related dimensions of sustainability. It is advocated that those involved in operationalizing sustainability (especially in the context of creating post-2015 Sustainable Development Goals), should explore the potential for developing indicators to capture some of its less tangible aspects, especially those concerned with ethical values.

Chapter 3. Operationalising the 'missing pillar' of sustainability: towards intersubjective values-based indicators

3.1. Introduction: what have we all been missing?

3.1.1. The "Missing Pillar" of Sustainability: A Convergence of Perspectives

There is a significant growing concern in several arenas that the 'three-pillar' model of sustainability, consisting of environmental, economic and social dimensions, may be overlooking something of fundamental importance. As highlighted by Littig and Griessler (2005) and more recently by Dahl (2012), there have been several attempts to define this missing dimension as a fourth pillar of sustainability, but it has been variously described as a *cultural-aesthetic*, *political-institutional*, or *religious-spiritual* dimension:

Cultural-aesthetic. A well-established framing of the fourth pillar, or missing dimension, of sustainability conceptualizes it in terms of culture, the arts and/or aesthetics. Jon Hawkes makes this case explicitly in his book *The Fourth Pillar of Sustainability: Culture's Essential Role in Public Planning*, where he argues that 'cultural vitality', understood in the sense of 'wellbeing, creativity, diversity and innovation', should be treated as one of the basic requirements of a healthy society (Hawkes, 2001, p. 25). While advocating for community involvement in arts practice, Hawkes makes it clear that he is referring to a broader definition of culture that is not limited to arts and heritage, but encompasses the 'whole complex of distinctive spiritual, material, intellectual and emotional features that characterize a society or social group', as outlined in the 1982 Mexico City Declaration on Cultural Policies (UNESCO, 1982).

UNESCO has similarly been active in promoting the cultural perspective, and many of its publications since the 1990s have highlighted the central role of culture in sustainability— either as a 'self-standing pillar of sustainable development' (e.g. UNESCO, 2013, p. 7) or as a foundation underlying the other three pillars (De Leo, 2012). This has been particularly significant within the context of Education for Sustainable Development (ESD), where the

cultural pillar has a strong focus on acknowledging and respecting diverse worldviews, identities and local languages and promoting open dialogue and debate. As discussed by Sacha Kagan in *Art and Sustainability* (Kagan, 2011), there have also been a number of international declarations and processes aimed simultaneously at raising awareness of sustainability within the arts and culture sectors, and at incorporating a cultural-aesthetic dimension into ongoing sustainability discourses. These include, among others, the 2001 'Tutzinger Manifest' (a call issued at a conference on Aesthetics of Sustainability in Tutzing, Germany, for a cultural dimension to be integrated into the Agenda 21 processes arising from the 1992 Rio Earth Summit); the 'Agenda 21 for Culture' initiative led by the United Cities and Local Governments initiative, a coalition of local governments from different countries emerging from the 2004 Forum of Local Authorities for Social Inclusion; and the 'Culture Futures' conference co-organized in 2009 by prominent civil society organizations in the arts and culture sector, as a parallel to the COP15 UN Climate Conference in Copenhagen. The latter conference has generated ongoing activities and collaborations among the organizations concerned, such as the 'Connect2Culture' initiative of the Asia-Europe Foundation.

A 'cultural-aesthetic' perspective can also be found among Indigenous communities and their advocates, including the UN Food and Agriculture Organization (FAO), who frame the missing-pillar debate in terms of 'cultural integrity'. This term is used to encompass shared values, beliefs and knowledge, as well as more tangible manifestations of culture such as ceremonies and objects (Nurse, 2006; Woodley, 2006). Within this context, the United Nations Permanent Forum on Indigenous Issues has acknowledged the need for culturally appropriate indicators of well-being and sustainability that reflect 'true indigenous perspectives such as portraying approaches grounded in wholism [sic] and unique values' (UN-PFII, 2006, p. 7). As recently highlighted by Barkin and Lemus, these Indigenous understandings of well-being and sustainability may be rooted in epistemologies that are very different from those of mainstream sustainability discourses.

Political-institutional. The concept of a 'political-institutional' fourth pillar is also widely known. Institutional aspects of sustainability were explicitly addressed in the indicator system developed by the Commission on Sustainable Development (CSD) in 1995 to assess implementation of Agenda 21 (Pfahl, 2005; Spangenberg et al., 2002) as well as being the

subject of a dedicated chapter in the Brundtland report, *Our Common Future* (World Commission on Environment and Development, 1987). As Spangenberg explains, institutions are 'the result of interpersonal processes, such as communication and co-operation, resulting in information and systems of rules governing the interaction of members of a society' (Spangenberg, 2002, p. 104). The development of institutional sustainability indicators is rooted in an understanding of institutions which includes, but is not limited to, organizations: it also encompasses two other categories, namely *institutional orientations* (norms) and *institutional mechanisms* (formal systems of rules and procedures, whether administrative, social, political or legal) (Spangenberg, 2002; Spangenberg et al., 2002).

Beyond the initial Agenda 21 context, the use of the institutional dimension as a fourth pillar of sustainability has gained widespread acceptance within the European Commission and the United Nations. The System of Environmental and Economic Accounting (SEEA) refers directly to 'the three-pillars approach (with sometimes a fourth—institutional—pillar)' (European Commission: EUROSTAT, p.5). The United Nations Division for Sustainable Development also incorporates institutional indicators into its framework of sustainable development indicators (UN Division for Sustainable Development. Department of Policy Co-ordination and Sustainable Development, 1995, 1996; UN Division for Sustainable Development. UN Department of Economic and Social Affairs, 2000).

Religious-spiritual. A third, and much lesser-known, perspective on the missing pillar/dimension of sustainability is rooted in the concept of an awakening global ethical and spiritual consciousness that underpins sustainability transitions (Clugston, 2011; ECI Secretariat, 2010; Hedlund-de Witt, 2011). In his keynote address at the 2010 Earth Charter conference 'An Ethical Framework for a Sustainable World', Steven Rockefeller described this emerging consciousness as 'in truth the *first* pillar of a sustainable way of life', on the grounds that ethical vision and moral courage are essential to generating the political will required for transitions to sustainability (Clugston, 2011, p. 174, emphasis added). A similar sentiment is expressed through a slightly different metaphor in the Interreligious Statement to Rio + 20 (2012), developed by religious and spiritual leaders from diverse traditions, which describes ethical/spiritual consciousness as 'the foundation of the other three pillars'.

While these three conceptualizations of the missing pillar of sustainability may appear disparate at first sight, we propose that one thing which they all have in common is a concern

with *human values* and how they are manifested in people's personal and professional lives (Clugston, 2011; Pfahl, 2005; Woodley, 2006). This is not, of course, intended to imply that the dimension of values covers the entire scope of the above perspectives on the fourth pillar of sustainability, as they all encompass multiple constructs which are interrelated in complex ways. Nonetheless, we suggest that values constitute an important and hitherto unrecognized area of common ground between the perspectives, and that there is a strong case for highlighting values as a key element of the less tangible dimension that tends to be omitted from international sustainability discourses.

The term 'values' conceals a multiplicity of contested and often conflicting meanings, deriving from many different disciplines of academic research (ranging from moral philosophy to empirical social psychology), and discussions of values span the epistemological divide between the natural and social sciences and the humanities. For the purposes of this paper, however, we have found the twofold definition provided by the Oxford English Dictionary to be helpful for resolving some of the confusion that often surrounds its everyday usage. The dictionary defines values both as 'principles or standards of behaviour' (Definition A) and as 'one's judgement of what is important in life' (Definition B) (Oxford English Dictionary, 2013). We would argue that Definition A encompasses Definition B, in that the creation of principles or standards for the ways in which people 'ought to behave' is inherently rooted in judgements about what is important in life, whether or not these judgements are explicitly articulated. It is the acceptance of a certain world-view, or set of life priorities, that generates the sense of 'oughtness'—the compulsion to choose an apparently greater good over an apparently lesser good, or an apparently lesser evil over an apparently greater evil (Bahm, 1994). We note that neither definition requires the use of specific 'value-labels' (such as trust, justice, collaboration, *etc.*): many aspects of our lives are not directly linked to specific, conceptualized values but can still be values-based, e.g., the education we choose for our children.

Values in the sense of Definition B (people's judgments about what is important in life) have long been recognized as one of the major transformative forces influencing the ability to fulfil human needs (Berg & Nycander, 1997). It has been argued, in this sense, that even natural sciences can never be value-neutral (Lele & Norgaard, 1996) and that all sustainability indicators are therefore values-based to some extent, even though some (especially in the

environmental domain) attempt to hide their underlying values beneath a 'façade of objectivity' (Bell & Morse, 2008). We suggest, however, that what is absent from many discussions of sustainability is an explicit consideration of values in the sense of Definition A (principles or standards of behaviour), which we will henceforth refer to as 'ethical values' for the sake of clarity. Such ethical values can be viewed as linking discussions of cultural integrity and vitality (cultural/aesthetic perspective), personal growth (religious/spiritual perspective) and good governance (political/institutional perspective).

We acknowledge that the use of the 'fourth pillar' metaphor to signify the existence of a less tangible dimension to sustainability, with ethical values as one of its key elements, is an over-simplification—especially as ethical values permeate every field of human endeavour represented by the three existing pillars, and the social dimension in particular is intimately concerned with ethical values such as equity and justice (Littig & Griessler, 2005). We can envisage alternative metaphors, e.g., ethical values as a foundation or lintel for the three pillars, as a weft running through them, or as a spiral that winds around them—in each case touching each of them and linking them together, but also incorporating something extra. Our main reason for promoting the fourth pillar metaphor, instead of any other viable metaphor, is to advocate for the inclusion of ethical values in international sustainability discourses on equal terms with the three existing pillars. This, we suggest, could facilitate the mainstreaming of the concept of ethical values and reduce the likelihood of its being dismissed as a trivial detail.

3.1.2. Ethical Values in International Sustainability Discourses

The theme of ethical values as a crucial dimension of sustainability is taken up strongly in the Earth Charter (Clugston, 2011; Corcoran et al., 2005; ECI Secretariat, 2010) and the United Nations Millennium Declaration (United Nations General Assembly, 2000), which both call explicitly for an 'ethical framework' for sustainability and list specific values which they associate with it. The text of these documents, quoted in Table 3-1 below, makes it clear that they are referring to values in the sense of ethical principles or standards for behaviour—albeit underpinned by shared understandings of what is important in life. The Millennium Declaration also makes the first attempt to provide definitions of some of these ESD-related ethical values,

albeit in a very general way. A similar perspective is evident in the United Nations Educational, Scientific and Cultural Organization (UNESCO) International Implementation Scheme for the 2005–2014 Decade of Education for Sustainable Development (DESD), where both of the above meanings are clearly evident (De Leo, 2012). This document states, for example, that 'the basic vision for the DESD is a world where everyone has the opportunity to benefit from education and learn the values, behaviour and lifestyles required for a sustainable future and for positive societal transformation' (UNESCO, 2004).   De Leo (2012) has conducted a content analysis of 22 international documents relevant to ESD, dating from 1945 to 2006 inclusive, and identified from them 16 frequently mentioned 'global values' (along with 17 less frequent ones). All of the frequently mentioned values, except 'freedom', are also mentioned in the International Implementation Scheme for the DESD.

By contrast, in the 2002 Rio + 10 report (United Nations, 2002) the word `values' is mentioned only once, in a small section reporting on a roundtable that refers specifically to the values of democracy as freedom, equality, tolerance, and respect for nature. In the final outcome document from Rio + 20, The Future We Want (United Nations, 2012), the concept of values (either in the sense of ethical principles or of value judgments) disappears altogether. The Rio + 20 document does not explicitly mention the word 'values' at all, except in a third sense relating to worth (ecological, economic and other 'values of biological diversity').

It could be argued that the momentum generated around ethical values in the international sustainability discourse at the turn of the millennium has already been lost. An alternative interpretation is that the Commission for Sustainable Development (CSD) process and the Millennium Declaration/Earth Charter process actually represent parallel discourses. Koroneos and Rokos (2012) suggest that the paradigm of 'sustainable development' espoused by the CSD process has already been co-opted by the neoliberal economic growth agenda, and outline an alternative paradigm of development that is fundamentally rooted in ethics and human well-being, which they term '(worth-living) integrated development'. This latter paradigm may be more compatible with the understanding espoused in the Millennium Declaration and Earth Charter, as well as the ethical values-based position on ESD that has been adopted by UNESCO.

Whatever the underlying reasons, the lack of any reference to ethical values or principles in the Rio + 20 outcome document is a cause for great concern. Since the Rio + 20 outcome

document is clearly intended to guide the creation of the forthcoming Sustainable Development Goals (SDG), there is a strong likelihood that reference to ethical principles of sustainable behaviour will be entirely omitted from the SDG process, unless early and decisive action is taken by the relevant institutions to identify appropriate ethical values-related goals, targets and indicators.

Table 3-1. 'Values' in international sustainability/sustainable development documents, 2000-2012.

Source	Values of Sustainability/Sustainable Development (SD)	How described
Earth Charter	Respect and Care for the Community of Life Ecological Integrity Social and Economic Justice Democracy, Non-Violence and Peace	"We urgently need a shared vision of basic values to provide an ethical foundation for the emerging world community" [29] (p. 1) "…we affirm the following interdependent principles for a sustainable way of life as a common standard by which the conduct of all individuals, organizations, businesses, governments, and transnational institutions is to be guided and assessed" [29] (p. 1)
UN Millennium Declaration (2000)	Freedom, Equality Solidarity, Tolerance Respect for Nature Shared Responsibility	"We consider certain fundamental values to be essential to international relations in the twenty-first century." [30] (Section 6)
"Global values" (1945–2006)	Equality (561) Responsibility (474) Participation (455) Cooperation (292) Dignity (285) Freedom (279) Security (278) Peace (267) Protect (262) Respect (219) Dialogue (216) Integrity (189) Diversity (182) Tolerance (158) Justice (134) Solidarity (85)	This list contains the 16 most frequent "global values" identified through a content analysis of 22 international documents issued between 1945 and 2006 inclusive [6]. Each word in the list represents a cluster of similar concepts or meanings (e.g., "Peace" represents "peace, non-violence, harmony, social cohesion"). The numbers in brackets refer to the number of times that the respective value cluster was mentioned in the 22 documents. Less frequently mentioned values were love, nurturing, privacy, confidentiality, informed/free consent, innovation, creativity, imagination, empowerment, generosity, gratitude, humility, wisdom, resilience, hope, stability, and reverence (*i.e.*, for life and the environment). These values do not appear as lists in the original documents; their collation into list form has been done by De Leo.
Final Report from Rio + 10 (2002)	Freedom Equality Tolerance Respect for Nature	"Some participants [in Round Table 3] highlighted links among environment, trade, peace accords, military arms reduction, the implementation of the Monterrey agreement and developing the values of democracy within a sustainable development framework. They supported adherence to the principles…" [32] (p. 127)
Final Report from Rio + 20 (2012)	None listed	"We reaffirm the intrinsic value of biological diversity, as well as the ecological, genetic, social, economic, scientific, educational, cultural, recreational and aesthetic values of biological diversity…" [33] (p. 38)

In this paper, we aim to contribute to ongoing conversations around the nature and roles of post-2015 Sustainable Development Goals by examining a logical argument for the exclusion of ethical values from sustainability assessment processes (the 'immeasurability' argument); illustrating that it is based on a false premise; and replacing it with an alternative logical argument (the 'context-specific measurability' argument). We also discuss practical considerations relating to the introduction of values-based indicators, and provide an illustrative example of their use.

3.1.3. Toward Ethical Values-Based Indicators

Existing frameworks of sustainability indicators specifically acknowledge the importance of 'values' in the sense of shared priorities (our Definition B). The European Reference Framework for Sustainable Cities (RFSC), for example, refers directly to the importance of building an integrated vision for sustainable development based on clear priorities and objectives, identified through processes of consultation with multiple stakeholders. Indicators that explicitly operationalize ethical values, in the sense of measuring the enactment of widely accepted principles or standards of behaviour, are, however, currently lacking.

In the light of the continuing high-level process to develop a global set of post-2015 Sustainable Development Goals, it is timely to ask why, given the widespread recognition of the importance of ethical values for sustainability transitions, the ethical dimension has so often been overlooked in the development of goals and indicators. Even in the arena of institutional sustainability, where the question of 'good governance' encompasses many aspects that could be considered values-based, efforts to develop indicators have been fraught with problems. In an assessment of Agenda 21 implementation conducted by the Commission on Sustainable Development, for example, a full 60% of the proposed institutional indicators were dropped before the final version, leaving the institutional dimension with only a third as many indicators as each of the other three dimensions. Many of the indicators dropped from the initial draft had been perceived during field testing to be unclear, irrelevant, lacking data, or in need of further development (UN Department of Economic and Social Affairs, 1999). Those indicators that remained did not adequately address the values aspect of institutions: Spangenberg (2002) criticizes the Agenda 21 assessment process for using an incomplete definition of 'institutions', often treating the term as synonymous with 'organizations', and not going far enough to operationalize less-tangible institutions such as values, norms and informal rules.

The omission of some draft indicators of institutional sustainability because of challenges with data availability (Spangenberg, 2002; UN Department of Economic and Social Affairs, 1999) hints at a broader concern about contemporary sustainability assessment. It might be assumed by a lay person that indicator development would precede data collection, but as indicator specialists will attest, the reverse is often true. As noted by McCool and Stankey (2004, p. 295), in the absence of broad public debate about what constitutes sustainability, efforts to develop new indicators are 'guided more by what *can* be measured (a technical issue) than by what *should* be measured (a normative issue)'. Indicator development thus appears to be severely constrained, whether by a failure of imagination, resource provision, or both: no matter how important something might be to the public, if it is not currently measured for other purposes (e.g., government statistics) there may be less interest in exploring whether it might be measurable.

Concerns about data availability do not, however, fully explain the lack of investment in values-based sustainability indicators. Another possible reason is given by Dahl (2012, p. 16) when he explicitly asserts, citing the seminal work of Hitlin and Piliavin (2004), that no values-based indicators have yet been developed because values are 'difficult to define and measure, with few widely accepted or standardized methodologies'. In lay discourse and a large body of academic literature, especially in the humanities, values tend to be characterized as highly subjective, subconscious, intangible, affective, context-bound and/or dynamic in nature, with an underlying assumption that there can never be any scientifically valid way of 'measuring' them. Redclift and Benton (1994), for example, state that people's values are 'negotiated, transitory and sometimes contradictory', while the popular author Daniel Goleman (1998, p. 57) powerfully conveys the affective nature of personal values by describing them as 'not lofty abstractions, but intimate credos that we may never quite articulate in words so much as *feel*' (emphasis in original). Meglino and Ravlin (1998, p. 360) similarly refer to values as being —less than totally conscious, somewhat below an individual's level of complete awareness‖. These widely shared subjectivist ontological assumptions about values appear to bring them into direct conflict with concepts of measurement: one cannot envisage measuring something which can neither be unambiguously conceptualized, nor adequately operationalized (Schlater & Sontag, 1994; Stapleton & Garrod, 2008).

Does it matter, then, if there are no useful indicators for internationally advocated ethical values such as those listed in Table 3-1? We would argue that it matters greatly, because indicators often have conceptual and symbolic uses far beyond their instrumental uses (Grainger,

2012; Gudmundsson, 2003; Rosenström, 2006): they do not merely reflect what is important in society, but also contribute to *defining* what is perceived as important (Baha'i International Community, 1998; Meadows, 1998). The current absence of ethical values-based indicators may contribute to perpetuating a situation in which 'nation-states are "managed" to enhance GDP ahead of almost all other concerns' (Bell & Morse, 2011, p. 225), while thorny issues of moral accountability are neatly evaded. Conversely, it might be expected that if values-based dimensions of sustainability were to be systematically assessed, one result might be the creation of new political norms (c.f. Rametsteiner, Puelzl, Alkan-Olsson, & Frederiksen, 2011) that tend to prioritize values such as equity, tolerance, justice and respect for nature at global, national and local levels.

In the next section, we explore the theoretical grounding for developing useful sustainability indicators based on ethical values, which can help to clarify some elements of the missing pillar of sustainability and render them measurable. We first distinguish between values espousal, which is widely measured through survey instruments in the empirical tradition of social psychology, and values enactment, which has not yet been systematically operationalized. Focusing on values enactment, we then outline a logical argument for excluding it from sustainability assessment discourses, and illustrate that this argument is based on a false premise. In Section 3 we outline an alternative conceptualization focused on the development of intersubjectively valid (rather than objectively valid) values-based indicators for specific practical contexts, and highlight the immediate and significant implications for the SDG process. Finally, in Section 4, we recommend initial steps that could be taken immediately to bring this missing dimension into the SDG agenda.

3.2. Are Values Measurable?

3.2.1. What Do We Mean by "Values"? Espousal Versus Enactment

The suggestion that values are too intangible to be measured might be strongly disputed by researchers in the empirical tradition of social psychology, where values survey instruments (addressing values in the explicit sense of 'what matters to people', but also implicitly ethical principles and standards) have been developed and widely used for many decades. Rokeach, for example, in his widely cited 'Rokeach Value Survey' (RVS), claims to represent the entire domain

of human values in two separate lists of 18 words or short phrases representing, respectively, modes of behaviour and idealized conditions (`end-states') of existence. Respondents rank each set of items in order of importance as principles that guide their lives, and the resulting rankings are interpreted as indicators of the individual's personal values (Rokeach, 1973, 1979b). The Schwartz Values Survey (SVS) is broadly similar, but uses 56 predefined value items and replaces the ranking activity with a rating scale (S.H. Schwartz, 1992, 1994, 2007; S. H. Schwartz & Bilsky, 1987; S. H. Schwartz & W. Bilsky, 1990; S.H. Schwartz & W. Bilsky, 1990; S.H. Schwartz et al., 1999; S.H. Schwartz et al., 2001). There are numerous other survey approaches which, while differing in the detail of their operationalization, share the assumption that an individual's personal values exist as discrete verifiable constructs which can be deduced from their responses to a questionnaire (Allport et al., 1951; AMA, 2002; Davidov, 2010; Davidov, Schmidt, & Schwartz, 2008; House, Hanges, Javidan, Dorfman, & Gupta, 2004; Inglehart & Baker, 2000; Peterson, Park, & Seligman, 2005).

To understand why values surveys in the Rokeach/Schwartz tradition might not be sufficient to solve the problem of values-based sustainability indicators, it is important to appreciate that values can manifest themselves both through discourse (what people say) and overt action (what they do). In this paper, we follow Gruys et al. (2003) in using the term 'values espousal' to refer to the use of values-related vocabulary in spoken, written and/or multimedia discourses, and 'values enactment' to describe situations in which values are 'lived out or enacted [by individuals] through their specific actions and behaviours'—drawing on the earlier work of Argyris and Schön (1978)[66]. This formulation of 'values enactment' and 'values espousal' is derived from research in social psychology, but has parallels with work carried out in other disciplines, such as sociology—notably Bernard Lahire's duality of 'dispositions to act' and 'dispositions to believe' (Lahire, 2003), which in turn draws on the work of Bourdieu (1991).

As noted by Schlater and Sontag (1994, p. 5), there is often a mismatch at the individual level between the public espousal of values in discourse and their enactment in behaviour: 'A person may "talk" the value but not implement it in action, or a person may act in accordance with a value but not subscribe to it verbally.' Values surveys cannot measure enactment of values by individuals, organizations or states, but only what they are willing to articulate verbally: they do not offer any way of identifying whether there is a mismatch between the values implied by respondents' survey responses and those manifested in their real-life actions (Figure 2-1). We are not suggesting that what people say about their values is unimportant, but rather that attempts

to develop useful values-based sustainability indicators should examine the question of enactment (e.g., by observing behaviour or conducting surveys of peers) instead of relying entirely on self-report surveys.

Figure 2-1. The fallibility of self-report surveys for values

3.2.2. The Theoretical Possibility of Measuring Values Enactment

We turn now to the theoretical question of whether values *enactment* can ever be measured in a meaningful way through the use of indicators. To do this, we critically examine a logical argument for the immeasurability of values enactment, which might be used implicitly to block investment in the exploration of values-based indicators. We first present the argument in full, then investigate the truth of each of its premises, and finally determine whether the logical reasoning leading from the premises to the conclusion is sound (This analytical approach is adapted from Thwink.org).

The first premise, as implied by Dahl (2012) in his comment that values are 'difficult to define', is that obtaining an objective definition of values enactment, or of the enactment of a specific value, is impossible see also (Goleman, 1998; Hitlin & Piliavin, 2004; Redclift & Benton, 1994). We define 'objective', for this purpose, as existing independently of individual human understandings. The second, a basic assumption underlying indicator development processes, is that an objective definition is required in order to operationalize a concept as measurable indicators (Hinkel, 2011). As illustrated in Figure 2-2, these two premises taken together lead

logically to the conclusion that values enactment cannot be operationalized as measurable indicators—and therefore cannot be measured.

Figure 2-2. A logical argument that values enactment cannot be measured (the 'immeasurability argument')
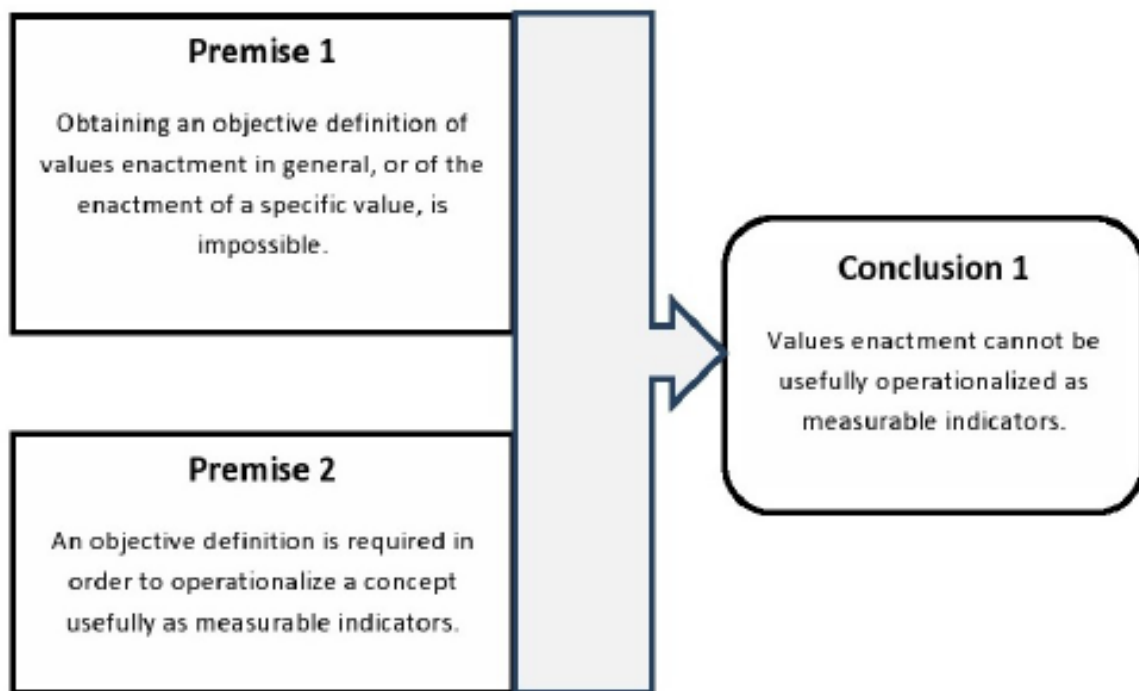
3.2.2.1. Can Values Enactment be Objectively Defined?

Earl Babbie (2010, p. 128) argues that values such as 'compassion' can neither be objectively defined (Premise 1), nor objectively measured. He illustrates this by using the concept of a 'mental file sheet' to represent an individual's conception of how a value might be enacted. If, for example, researcher X observes a subject ('Pat') performing certain actions that X associates with the word 'compassionate' (i.e., actions that appear on X's own mental file sheet for 'compassionate', such as putting a little bird back into its nest after finding it on the ground, or visiting a children's hospital at Christmas), she will conclude that Pat is compassionate. If, on the other hand, researcher Y does not observe in Pat's behaviour any of the actions that appear on his own 'compassionate' file sheet, and also notices Pat performing other actions that he regards

as indicative of a lack of compassion (e.g., refusing to donate money to a campaign to save whales from extinction), he will conclude that Pat is not compassionate. Babbie then comments:

> "We can measure, for example, whether Pat actually puts the little bird back in its nest, visits the hospital on Christmas…or refuses to contribute to saving the whales. All of those behaviours exist, so we can measure them. But is Pat really compassionate? We can't answer that question: we can't measure compassion in any objective sense, because compassion doesn't exist in the sense that those things I just described exist."

Further evidence from cross-cultural values studies supports the truth of Premise 1, that no objective definition of values enactment is possible, because people's personal understandings of values-related words and phrases ('value-labels'[11]) are heavily influenced by both their cultural contexts and their particular life experiences (Machicado & Davis, 1988; Peng et al., 1997; Torpe & Lolle, 2010). This point is similarly illustrated by literature in management and organization studies, where differences in the ways in which 'value-labels' are understood by managers and employees can contribute to significant problems within organizations (Cha & Edmondson, 2006; Gruys et al., 2008; Lencioni, 2002)—even leading directly, in one case, to the resignation of a senior executive (Lencioni, 2002). We therefore conclude that Premise 1 is true and that values enactment cannot be objectively defined.

3.2.2.2. Is Objective Definition a Prerequisite for Operationalization and Measurement?

The quest for indicators is, at first glance, inherently objectivist. The very word 'indicator' hints at its realist ontological assumptions: namely that there is an underlying 'theoretical variable' that exists in objective reality, and that its presence can be unambiguously indicated by one or more, similarly existent, 'observable variables' (Hinkel, 2011). Premise 2 is entirely consistent with this paradigm of indicator development, but to accept the premise as true is to deny the possibility of any alternative way of understanding indicators.

In the social sciences, as highlighted by Babbie (2010), 'most of the variables we want to study don't exist in the way that rocks exist. Instead, they are made up. Moreover, they seldom have a single, unambiguous meaning.' This does not, however, mean that nothing can ever be operationalized or measured. Rather, useful constructs can be created through mutual agreement

---

[11] The term 'value-labels' was created by Marie Harder during the ESDinds project (ESDinds, 2011).

for the purpose of communication and research—a process termed 'conceptualization'. What this entails is not objectivity but intersubjectivity, a complex and multi-layered concept that we might summarize for the purposes of this argument as 'the emergence of a human "interworld" of shared meaning that transcends individual consciousness (Crossley, 1996, p. 4). This shared meaning emerges in a collaborative context through dialogical interactions grounded in a common lived experience, such as a jointly undertaken practical activity (Talamo & Pozzi, 2011).

To extend Babbie's earlier example (c.f. Babbie, 2010): if the two researchers in question were employed in a teaching hospital that explicitly aimed to train nurses to be compassionate towards patients, they might already have a degree of shared understanding about what a 'compassionate' nurse should do (or avoid doing). Thus, their 'mental file sheets', rather than being completely different, would overlap to a certain extent. Through a formal process of dialogue with hospital stakeholders, they could translate this informal shared understanding into specific measurable indicators and assessment tools, and hence evaluate the extent to which 'Pat'—as a final-year student nurse—is enacting the value of compassion in the sense desired by the hospital. It is true that these particular indicators of compassion are unlikely to be generalizable universally, and that they would not encompass every possible aspect of enacting compassion. There is no doubt, however, that they could be locally well-defined; and one can envision that they might provide the hospital stakeholders with useful information about the success or failure of Pat's training.

We can thus conclude that Premise 2 is false, i.e., an objective definition of a value such as 'compassion' or 'respect for nature' is not a prerequisite for creating useful indicators of its enactment. Indeed, much of the measurement-based work conducted in the social sciences negates Premise 2 (Babbie, 2010). Rather, what is important is that the value is intersubjectively defined in relation to a specific, bounded practical context (Talamo & Pozzi, 2011) and that the intersubjective definition is accepted by the individuals involved as a useful, if incomplete, working definition.

This implies that 'values enactment' is measurable within a specified context, provided that certain parameters are predefined, and leads us to propose the replacement of the 'immeasurability' argument with an alternative logical argument (which we term the 'context-specific measurability' argument) that opens the way for the development of values-based sustainability indicators. This argument is summarized in Figure 2-3, below. In the following

section, we briefly present an illustrative example from our own work in support of this theoretical argument.

Figure 2-3. The 'context-specific measurability' argument
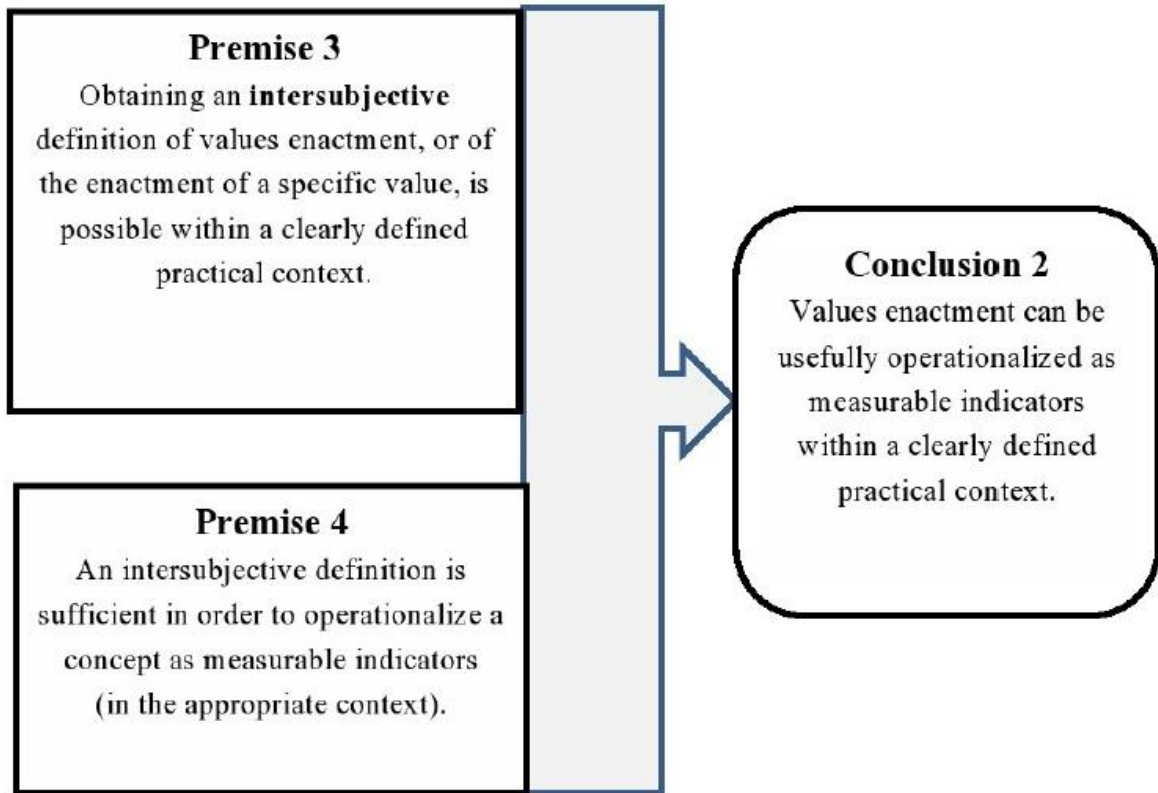
3.3. Developing and Using Values-Based Indicators: An Illustrative Example

We have asserted that through a process of intersubjective conceptualization within a clearly defined context, the enactment of human values (previously regarded as intangible) can be operationalized and measured. In this section, we will provide a practical example, derived from a collaborative international research project, to illustrate how processes of conceptualization and operationalization of values-related dimensions of sustainability have been used to develop project-level indicators.

3.3.1. Background: The ESDinds Project

The ESDinds project, which aimed to develop values-based indicators and assessment tools for civil society organizations promoting education for sustainable development, was funded by the European Union Seventh Framework Programme (FP7) from 2009–2011 [79–83]. The project brought together representatives of two academic research institutions and four civil

100

society organizations (CSOs) as equal partners, and was innovative in the extent of decision-making power granted to the CSO partners within the research consortium.

The first phase, which we termed values elicitation, comprised the identification of a pool of value-labels and related pilot indicators ('proto-indicators') through content analysis of a large qualitative dataset generated through case study research, key informant interviews, workshops and document analysis within six 'source' CSOs (Podger et al., 2010). As the initial data analysis generated a very long list of values that the CSOs regarded as important for their work, prioritization was carried out on the basis of coding frequency to generate a list of five values with their associated proto-indicators: integrity, trustworthiness, unity in diversity, empowerment and justice. A sixth value, care and respect for the community of life, was also added after two members of the consortium objected that the overall set of values was incomplete without it. A total of 177 draft proto-indicators were intersubjectively chosen by representatives from all source CSOs, from the much larger number extracted from the data set for these six values (ESDinds, 2011).

In the second phase, exploratory field work, the peer-elicited proto-indicators were field-tested with 'user' CSO partners acting as 'critical friends' (Podger et al., 2013). Extensive testing of the indicators was conducted in 15 organizations, each of which selected between 3 and 25 indicators to measure locally (ESDinds, 2010a, 2011; Podger et al., 2013). The feedback was used to improve the relevance/importance, validity, comprehensibility and measurability/usability of the indicators in a variety of different civil society contexts, in accordance with current recommendations for developing sustainability indicators (Reed, Fraser, & Dougill, 2006). Through this process, the indicators were discovered to be very broadly relevant across diverse organizational and cultural contexts, and to be a sufficiently large pool to construct indicator sets for several other common human values of importance to CSOs (ESDinds, 2011). Another important finding was that the usefulness of the indicators was greatly enhanced by localization, i.e., the flexibility to modify the wording to suit the specific practical context: although this also has the disadvantage of precluding direct comparisons between different organizations, it had the advantage of providing excellent face validity. Furthermore, generalizability was not entirely lost as each indicator 'item' could still have measures compared qualitatively across organizations or time or circumstances. Once an intersubjective local consensus was reached for those indicators chosen to be of priority locally, devising means of obtaining measures of them became

a feasible task. That is, the localization provided the boundary conditions which allowed clear specifications for useful measures to be devised.

One example of the use of this approach is given below in Section 3.2, where we illustrate the use of values-based indicators developed through the ESDinds project to evaluate an online course in sustainability leadership. A separate evaluation of the ESDinds method for purposes of project monitoring and evaluation in three different organizations is described in our earlier work (Burford, Velasco, et al., 2013), and there are many other possible applications for these indicators at different levels and in different contexts. The purpose of the case study in this specific paper is to demonstrate the fallacy of the immeasurability argument for enacted values.

3.3.2. Measuring Values Enactment at the Project Level

One example of the many applications for the indicators developed during the ESDinds project was a values-focused evaluation for a semester-long online course in sustainability leadership. The course seeks to train a new generation of sustainability leaders, with the specific mandate to 'promote values in relation to a shared ethics for sustainability' and build capacity in project planning, management, fundraising and social media utilisation (ECI Secretariat, 2011). The aim of using the ESDinds toolkit to evaluate the course was to assess the extent to which specific ethical/spiritual values, associated with sustainability leadership and the Earth Charter, were present in (a) the course content; (b) the course implementation; and (c) participants' behaviour during and, where possible, after the course.

The first step in this evaluation was a workshop exercise in which the course facilitators intersubjectively identified and defined which of their values they wanted to assess the presence of, using their own local terminology. In the second step, facilitators read through the full reference list of ESDinds indicators and selected 49 indicators that they felt represented their locally defined value concepts and were highly relevant to the course, without concerning themselves about how they might be assessed. In the third step, the group re-read the indicators again as a set, reduced the list to 40 indicators on the basis that some of them were felt to be duplicated, and reflected on which assessment methods might be appropriate.

A group consultation approach was then used to design a mixed-methods assessment strategy to provide measurements of the 40 indicators. The final chosen (localised) assessment methods were, first, developing a participant questionnaire with both open-ended and multiple-

choice questions; and second, conducting a qualitative content analysis of participants' feedback (obtained through the course's Facebook page and inbuilt feedback mechanism) and of their submitted assignments, which included proposals for sustainability projects in their home communities or elsewhere. These methods were locally deemed sufficient to provide 'measures' of local values, for the given context and purpose. The quantitative findings from the survey questionnaire were then converted to qualitative data, and the entire dataset was analysed in terms of participants' and/or facilitators' enactment of each of the five locally defined value clusters: (a) Empowerment; (b) Participation; (c) Innovative Thinking; (d) Equality, Inclusiveness, and Tolerance; and (e) Transformation for Sustainability and Universal Responsibility. The project coordinator felt that the evaluation findings provided the facilitator group with a deep qualitative understanding of how each of the above-mentioned values was enacted in the online course. They provided confirmation of its overall effectiveness in catalysing value change towards sustainability, and also highlighted several ways in which the course content or delivery could be improved in order to score more highly on specific values-based indicators.

While the small sample size for the online course meant that it was not necessary or desirable to quantify the data in this particular case, the intersubjective approach to values measurement does allow for quantification and statistical analysis where appropriate. In a larger study, for example, it would be possible to pre-test the questionnaire for internal consistency (e.g., Cronbach's alpha) and then intersubjectively determine benchmarks. A group might decide, for example, to rate an indicator as 'green/good' if more than 75% of responses to an indicator are positive (according to their own collectively agreed definition of what would constitute a positive response), 'amber/satisfactory' if 50–75% of responses are positive, and 'red/unsatisfactory' if less than 50% of responses are positive. The ratings for individual indicators could then be aggregated to give an overall rating for the value[12]. We are working on developing quantitative assessment tools for measuring specific pro-sustainability values within a higher education context[13]. The above example illustrates that it is possible to operationalize human values intersubjectively through dialogue, by building consensus around the linkage of locally defined value concepts to specific measurable indicators. These indicators may be drawn directly, and/or modified, from a generic list such as that generated by the ESDinds project. The local meaning of each value-label (e.g., 'empowerment' or 'equality') thus becomes defined by

---

[12] Subsequently published as Podger et al. (2016).
[13] Subsequently published as Ribeiro et al. (2016).

consensus-validated referents (behaviours, perceptions and aspirations) within the clearly stated context, in this case the online course. We are not claiming, of course, to have defined 'empowerment' in a universally applicable way, but only to have defined the type of empowerment that the facilitators of this specific course were seeking to achieve. However, the overlap between multiple intersubjective definitions of values enactment would be an interesting arena for future research, especially where the contexts are broadly similar. What this means is that named ethical and spiritual values can be 'measured' in a useful and locally valid way, through further dialogic processes of devising appropriate data collection and analysis strategies and establishing benchmarks where necessary. According to the requirements of the situation, these strategies may be qualitative, quantitative, or mixed in nature, and may involve a combination of standard methods (e.g., survey questionnaires) and innovative context-specific methods (in this case, content analysis of comments left on the intrinsic feedback mechanism built into the online course platform).

Our empirical findings therefore support the theoretical argument made in Section 3.2.2.2., namely that it is possible to obtain an intersubjective definition of values enactment within a defined local context (Premise 3) and that the intersubjective definition is sufficient to allow operationalization and measurement of values in that context (Premise 4). Since the measurements obtained in the case study were useful to the team of facilitators, we suggest that Conclusion 2 might be enhanced by the addition of the word 'usefully', thus: 'Values enactment can be usefully operationalized as measurable indicators within a clearly defined practical context'.

3.4. Discussion

3.4.1. Potential for the Further Development of Values-Based Indicators

In this paper, we have illustrated that there is no actual theoretical barrier to the measurement of ethical values, which constitute a key element of the 'missing pillar' of sustainability. Although objective and universally accepted definitions of values may be impossible to attain, we introduce work where a useful set of localizable values-based indicators has been successfully developed and utilized at the project level, through a process of intersubjective conceptualization. We propose, therefore, that it is no longer valid to argue that

values-based aspects of sustainable development cannot be usefully measured and that, in the light of its inherent flexibility, the ESDinds method could be adapted for developing values-based sustainability indicators in wider contexts. This has significant implications, not only for emerging transnational grassroots movements such as the Earth Charter Initiative (ECI Secretariat, 2010) that are keen to understand the strengths and weaknesses of their existing efforts, but also for the entire global apparatus dedicated to defining and measuring sustainability. Even if difficulties were to be found in adapting this exact method, its success thus far may be regarded as 'proof of concept' that indicators can be developed for less tangible constructs than might previously have been imagined.

We strongly recommend, therefore, that the institutions tasked with developing Sustainable Development Goals, targets and indicators should take time, at this critical juncture, to pause and reframe the sustainability assessment process. We echo McCool and Stankey (2004) in calling for a shift away from a convenience-driven technical approach ('what *can* be measured', using the methods and datasets that are currently available), towards a normative approach based on creative and critical thinking ('what *should* be measured'). In our view, what is required at this stage is not mere accumulation of indicators (Grainger, 2012), but greater efforts to conceptualize and operationalize 'sustainability' from scratch in diverse contexts where explicit discourses of ethical values already exist—including grassroots transnational networks such as the Earth Charter Initiative that promote a clear set of principles, faith-based organizations which articulate values, and Indigenous communities where customary laws incorporating shared values are known and respected.

While this process would almost certainly benefit from some degree of global coordination, it is imperative to avoid mistranslating our call for a 'global quest for values-based indicators' as a `quest for global values-based indicators'. Following Hulme (2010), we suggest that a one-size-fits-all approach on a global scale is unlikely to yield useful results, especially because intersubjective processes tend to be inherently rooted in local contexts of shared practical activity (Talamo & Pozzi, 2011). What may be more helpful is a polycentric approach, based on the creation of multiple, diverse, peer-elicited, indicator sets and assessment tools (Hulme, 2010; Ostrom, 2010). It might be useful, for example, to create reference sets of indicators and survey questionnaires which can be used in their standard forms to generate national-level statistics, but can also be tailored to local contexts by CSOs, local government institutions, religious groups, companies and Indigenous communities as part of an emerging

culture of self-evaluation, learning and change. Awareness of the need for such contextualization for practical purposes has been raised previously (Patton, 2007).

One possible approach would be to begin by developing multiple small-scale frameworks of values-based indicators, beginning at the project and organization levels, and then (a) expanding their reach vertically to neighbourhood, city, district, regional and national levels; and/or (b) expanding their reach horizontally via transnational grassroots networks of CSOs, such as the Earth Charter Initiative. In each of these cases, the initial exploration of shared values (in the sense of explicit or tacit principles/standards of behaviour applicable within the defined contexts) could be conducted through surveys or interviews that would be statistically representative of the respective populations. Subsequent stages of indicator development could then use approaches such as citizens' juries to check candidate indicators for perceived relevance, comprehensibility and usefulness. If there is initial resistance to the concept of values-based indicators on the part of decision-makers, one strand of research might initially focus on developing standardized assessment tools (e.g., survey instruments tested for internal consistency) for use at national and global levels, even while CSOs continue to explore flexible approaches tailored to their local realities. Such standardized tools might help to build confidence in the ability of values-based approaches to deliver rigorous evidence to inform decision-making, thereby providing a starting point from which more complex, participatory and mixed-methods approaches can subsequently be developed.

We accept that these suggested processes for developing values-based indicators are potentially complex and may pose implementation challenges, and that they contrast sharply with the inherent reductionism and top-down nature of many conventional processes for creating goals and indicators. The underlying principles of participation, co-design and including multiple stakeholder perspectives are not, however, without precedent. Combinations of 'top-down' and 'bottom-up' approaches to indicator development have already been employed by Reed, Fraser and colleagues, e.g., in the context of developing rangeland degradation indicators with indigenous herders in the Kalahari (Reed & Dougill, 2002; Reed et al., 2006; Reed, Fraser, Morse, & Dougill, 2005) and creating well-being assessments with stakeholders in coastal British Columbia (Fraser, Dougill, Mabee, Reed, & McAlpine, 2006). We firmly believe, furthermore, that the centrality of ethical values to sustainability—as attested by the Earth Charter and Millennium Declaration, as well as the work of UNESCO—justifies substantial investment in this area. This view is partially supported by the widespread enthusiasm for values-based indicators that has

been generated since the conclusion of the ESDinds project in December 2010. At the time of writing, the web platform created through the project has received over 8000 hits from 138 countries, and has generated an online community of interest with 143 members. Of these, 38% (n = 54) have engaged actively with the platform by marking (and, in many cases, customizing) those indicators that they find relevant to their work. The fact that only 35% (n = 50) of the total membership and 28% (n = 15) of those selecting indicators describe their affiliation as 'non-profit, charitable or humanitarian organization', with the others variously describing themselves as private companies (including social enterprises), faith-based organizations, public sector organizations, academic or educational institutions, families, informal community groups or individuals, attests to the usefulness and relevance of ESDinds values-based indicators beyond the specific sector in which they were developed (authors' unpublished results).

The political challenges of reframing sustainability to give due consideration to the enactment of ethical values should not be underestimated or trivialized. There are ongoing controversies surrounding the term 'sustainable development' in international discourse, including the critique of the implicit economic growth model and its relationship to powerful vested interests within the corporate sector (Bell & Morse, 2011; Koroneos & Rokos, 2012; Robinson, 2004). Those debates could be seen as existing in a separate plane, but generating parallel arguments that are highly relevant to the plane of sustainability assessment. We maintain, however, that at this critical time in history it is imperative to create a space—however loosely defined—within the Sustainable Development Goals process to allow for subsequent operationalization and measurement of ethical values. This would enable relevant research and practice to be supported, even while the broader and more complex questions of politics, ideologies, power relations and the appropriateness of the term —sustainable development‖ continue to be debated in multiple arenas. If this opportunity is missed now, it may become increasingly difficult for the voices of fourth-pillar advocates (of whatever persuasion) to be heard in the future.

3.4.2. Mitigating the Risk of Perverse Effects

Beyond the practical implementation issues that we have already discussed, another concern is the possibility of perverse effects arising from the introduction of values-based indicators. As in the case of other types of indicators, even though the goal of their introduction

is to increase rationality in decision-making, there may be some potential for forms of use which undermine this goal. Political use might, for example, lead to instrumentalization and data manipulation; while tactical use might result in the selective communication of results on the basis of personal interest (Krank, Wallbaum, & Grêt-Regamey, 2013). The flexible, localizable and largely qualitative nature of values-based indicators may render them particularly susceptible to misuse, and even if they are used as intended, care will need to be taken to avoid social desirability response bias when using assessment tools such as survey questionnaires or interview schedules (Arnold & Feldman, 1981; Fisher & Katz, 2000). One important feature of the ESDinds system is the use of mixed methods, ideally including at least one method that does not rely on self-report data, in order to reduce the overall effect of social desirability bias when measuring a particular indicator (as in the example above, where questionnaires were supplemented by a qualitative analysis of text that had already been submitted to an online platform). Furthermore, multiple linked indicators can be combined in order to provide information about a specific ethical value or cluster of values.

Within CSOs, a study of eight organizations where the ESDinds indicators have been used (including the example described above) identified a diverse range of positive effects, and no known negative effects (authors' unpublished results)[14]. It is possible, of course, that the respondents may have forgotten the perverse effects, chosen not to mention them, or failed to recognize them as attributable to the indicators. It is also possible that perverse effects took longer to become evident than positive effects, and had not yet emerged at the time of the study (3–6 months after the application of the indicators in each case). However, we suggest that in these organizations there may have been additional factors that contributed to the generation of positive rather than negative effects: (i) a shared understanding that the purpose of indicators is to assist local stakeholders to improve the effectiveness of the organization or project, rather than to enable external assessors to criticize its shortcomings; and (ii) a participatory approach in which local stakeholders, in this mode of learning, collaborate in identifying those indicators and assessment tools that they regard as relevant, important and interesting within their local context (ESDinds, 2011; Podger et al., 2010). Further research is necessary to determine whether any aspect of this learning is transferable to higher levels of indicator use, such as a neighbourhood, city or local authority.

---

[14] Subsequently published as Burford et al. (2016) and included above as Chapter 2 of this thesis.

3.4.3. Values-Based Indicators, Transdisciplinarity and Sustainability

The ESDinds research project described in this paper draws on a currently dispersed but vast body of work on values, participation and iterative co-learning. This work does not sit comfortably within either the 'objectivist' or 'subjectivist' paradigms, but rather, aims to promote interdisciplinary learning at the interface between the social and natural sciences (Conrad, 2002) in defiance of the apparent dichotomy between them. It could also be described as transdisciplinary research, i.e., research that looks not only across and between disciplines but also beyond them—at least in the sense in which transdisciplinarity is understood by authors such as Burger (2003) and Gibbons and colleagues (1994), as crossing the boundaries between 'science' and 'society'. According to this view, adopting a transdisciplinary approach means facilitating the deep participation of non-scientific stakeholders in research and the 'direct application of scientific knowledge in both political decision-making and societal problem-solving' (Burger & Kamber, 2003; Kagan, 2011). The focus of ESDinds was placed on societal problem-solving from the start, with the overarching criterion for decision-making at each point in the process being 'face validity'—the extent to which the emerging indicators or tools were regarded as relevant, important and interesting by the participating CSOs (Podger et al., 2013).

As noted by Basarab Nicolescu, however, the understanding of transdisciplinarity as joint problem solving at the science-society interface represents only one of a great many potential meanings of the term (Nicolescu, 2002, 2006, 2010). Nicolescu emphasizes that looking 'beyond disciplines' should not be limited to the social realm, and that complementary approaches (phenomenological, theoretical and experimental) are required if we are to understand those forms of knowledge that cut across and transcend diverse academic disciplines, as well as diverse cultures and religions (Nicolescu, 2006). Nicolescu's own theoretical insights, for example, include three axioms of transdisciplinarity supported by evidence from quantum physics: (i) the ontological axiom concerning the existence of different levels of Reality and different levels of perception; (ii) the logical axiom concerning the 'included middle', i.e., the possibility of being

simultaneously A and non-A; and (iii) the complexity axiom of universal interdependence (Nicolescu, 2002, 2006, 2010).

As the establishment of transdisciplinary approaches (in the broadest sense) to knowledge and education may arguably constitute a crucial element required for transitions to sustainability (Kagan, 2011; Nicolescu, 1998), the creation of Sustainable Development Goals could be a highly appropriate arena in which to examine the relevance of such ideas. In particular, an exploration of the topic of values-based indicators in relation to emerging theoretical understandings of transdisciplinarity may be highly informative, and we propose that it could offer potential for a profound transformation of the landscape of sustainability and indicators, although such an analysis is beyond the scope of this paper.

3.5. Conclusions and Recommendations

In this paper we have illustrated that it is theoretically and practically possible to assess processes and outcomes relating to the less tangible dimension, consisting of human values, ethics and worldviews, which we have chosen to conceptualize as an element of the fourth pillar of sustainability (while recognizing that alternative metaphors are also plausible). We thus strongly recommend that the institutions tasked with developing Sustainable Development Goals, targets and indicators should take time, especially at this critical juncture, to reframe the sustainability assessment process and incorporate an explicit acknowledgement of its ethical dimensions.

In particular, we recommend the following policy actions:

- Establishing a manageable but meaningful consultation process with key stakeholder groups within the institutions tasked with developing SDGs, to ensure that appropriate indicators and assessment tools relating to ethical values (as a key element of the 'missing pillar' of sustainability) are formulated in parallel with the goals themselves;
- Ensuring that projects initiated in support of SDGs are context-relevant and defined on the basis of consultation about local needs, priorities and values, rather than by the desire to improve national or global performance in relation to measurable indicators such as GDP; Facilitating or encouraging a funding mechanism to support rigorous research into indicators and assessment methodologies focusing on sustainability processes and outcomes that are less tangible, or more qualitative in nature, than those currently being

measured, including studies of perverse effects and how they might be mitigated or avoided;

- Using values-based indicators to reflect on some of the complex barriers to success in achieving the Millennium Development Goals (e.g., reframing conflict as a failure to enact values such as equality and tolerance, overexploitation of finite environmental resources as a deficiency in 'respect and care for the community of life', or systemic corruption as a lack of trustworthiness and integrity), in order to identify timely measures that might be taken to address these issues from a values perspective;

- Investing in research that addresses the issue of sustainability assessment in general, and values-based indicators in particular, through the lens of transdisciplinarity.

Chapter 4:

Can we improve indicator design for complex Sustainable Development Goals? A comparison of a values-based and conventional approach

Abstract

A conceptual framework was constructed for United Nations' complex Sustainable Development Goal (SDG) Target 4.7 focusing on education for sustainable development (ESD), and used to analyse the usefulness and character of indicators produced from a values-based approach called ESDinds, compared to a UN process. The analysis shows that the latter generated very few indicators concerning the wider aspects of knowledge such as 'critical thinking' or 'learning to learn'. The values-based approach, created for a different purpose, produced complementary if not better coverage of Target 4.7, including finely-developed concepts for competencies and less tangible aspects. It is suggested that the UN process would benefit from ESDinds design elements such as intersubjective and slightly disruptive elements, purposeful contextualisation at group level, and a holistic and inductive consideration of values. The use of a reference 'fuzzy framework' of slightly generalised proto-indicators suited for deep contextualisation locally is recommended, rather than any rigid global-level indicator with unclear local value. It is recommended that ESD practitioners immediately develop localised interpretations of valid measures for whatever final Target 4.7 indicator is selected by the UN, as this localisation process will itself cause important learning towards local ESD achievements.

Keywords:

Sustainable Development Goals; sustainability indicators; indicator design; values-based indicators; education for sustainable development; education for sustainability

Chapter 4. Can we improve indicator design for complex Sustainable Development Goals? A comparison of a values-based and conventional approach

4.1. Introduction: SDGs and their evaluation

The creation of the 2030 Sustainable Development Goals (SDGs) responded to the call by United-Nations Secretary-General Ban Ki-Moon for "the most inclusive and transparent development agenda the world has ever seen" (United Nations Millennium Campaign, 2015, p. 4). People in 194 countries contributed to the goals through a systematic process of 88 national consultations, 11 thematic dialogues, and a global online survey with over 8.5 million participants (United Nations Development Group, 2013; United Nations Millennium Campaign, 2015). The resulting goals are also distinctive in being applicable to all nations, regardless of gross domestic product or geographical location (United Nations, 2015b) which makes them a marked improvement on the Millennium Development Goals, which were widely criticised for 'ghettoising' development as an issue for the global South (Saith, 2006, p. 1184). The SDGs represent the first explicit acknowledgement, at the level of global goals, of the interconnectedness of the challenges surrounding sustainability and the corresponding need for integrated problem-solving (United Nations, 2015b).

While the SDGs are in many ways an important accomplishment, fundamental questions remain around how their implementation will be monitored and evaluated at all levels, from the global to the most immediately local. It is difficult to overstate the importance of identifying relevant and valid indicators. The 2015 Millennium Development Goals (MDGs) Report has explicitly acknowledged that "what gets measured gets done" (United Nations, 2015a, p. 10; see also Henshaw, 2006). The choice of indicators to measure progress towards the 169 SDG targets will contribute substantially to shaping national policies, strategies and action plans in most UN Member States, from now until 2030.

As noted by Gudmundsson (2003) and subsequent authors (Grainger, 2012; Hinkel, 2011; Rosenström, 2006; Rosenström & Kyllönen, 2007), the contributions made by indicators to sustainable development go far beyond their instrumental uses in providing data that may inform policy formation. In addition to their significance in shaping conduct, their mere

existence alters awareness, shapes attitudes and directs resources for the justification of decisions. These symbolic implications of indicators may be found entirely independent of, and even in conflict with, the empirical data which the indicators generate (Grainger, 2012). As an example, the use of gross domestic product (GDP) as a symbolic indicator of economic growth may be used to justify inaction or 'business-as-usual' policies that can directly block the achievement of ecologically-oriented goals (Bell & Morse, 2011).

The MDGs report calls for a "data revolution" through the use of the phrase "together we can measure what we treasure" (United Nations, 2015a, p. 13). This recognises the intimate relationship between goals, the values from which they spring, the indicators by which they are recognised, and the tools selected for their assessment. While deeply evocative, the use of the first-person plural in the statement 'we can measure what we treasure' implies singular identity and vision. This rhetorically compelling assertion erases the diversity of values and ideologies (at times incompatible) that characterise discussions bearing on sustainable development (SD) within institutions, whether global or local. Hopwood, for example, maps more than 30 different views on SD within a two-dimensional space—positioning them from 'low environmental concern' through 'techno-centred' to 'eco-centred' on the horizontal axis, and from low to high concern for socio-economic justice on the vertical axis (Hopwood, Mellor, & O'Brien, 2005). Thus although the need is acknowledged for indicators to accommodate plural viewpoints, to measure what 'matters' and to involve all those willing to participate, there is no clear pathway yet known to do this. The UN's approach may be the best under the circumstances, but it does not claim to be optimally designed.

In contemplating the many possible starting points towards designing effective working measures for the SDGs, the authors brought in a further dimension: the 'missing' fourth pillar. Without endorsing a particular definition of SD, we note assumptions and biases that the UN approach appears to perpetuate, which contribute to the marginalisation of those dimensions of sustainability that are not found within the established three dimensions of the economic, the environmental and the social. Any developments constrained to those three dimensions are insensitive, at minimum, to cultural/aesthetic dimensions, e.g., general discussions of cultural integrity and vitality (Hawkes, 2001; Nurse, 2006; UNESCO, 1982) and specific discussions in relation to Indigenous communities (UN-PFII, 2006); the role of the arts

in sustainability (Kagan, 2011); political/institutional dimensions, e.g., 'good governance' (Pfahl, 2005; Spangenberg, 2002); and religious/spiritual dimensions (Clugston, 2011; Hedlund-de Witt, 2011; Interreligious Statement Towards Rio+20, 2012). While in many ways these excluded dimensions are mutually distinct, they intersect in their shared interest in the category of values, in whatever manner these are understood. Although the domain of values certainly does not cover the full range of those excluded dimensions, it provides an entry point for bringing those dimensions into discussion with the three 'pillars' that appear to be more amenable to traditional forms of conception and measurement (see Chapter 3 above).

Values have historically been excluded from the realm of indicator development because they are perceived as intangible and difficult to measure (Dahl, 2012; Hitlin & Piliavin, 2004). However, it is possible to operationalise values when they are conceptualised intersubjectively (Chapter 3). That is, rather than viewing values as static latent traits possessed by individuals, they may become visible in groups when operationalised through systematic processes in clearly defined practical contexts, such as their normal working practice. An intersubjective approach to the design of indicators for values has recently been piloted in a variety of settings (Burford et al., 2015; Burford et al., 2014; Burford, Hoover, et al., 2016; Burford, Hoover, et al., 2013; Burford, Velasco, et al., 2013; Dahl et al., 2014a, 2014b, 2014c; Harder, Velasco, et al., 2014; Hoover et al., 2015; Podger et al., 2016; Podger et al., 2013; Ribeiro et al., 2016), including for formal evaluation. It typically involves an intersubjective process of values elicitation at the level of short statements about what is 'valuable, worthwhile and meaningful' to the group, e.g., "Mistakes are understood as opportunities to learn", coupled with gently disruptive probing to force conscious recognition of tacit knowledge, to denaturalise socially indicated responses, and finally to agree on local indicators. This inductive values-based approach to developing indicators differs from the conventional one used by the UN in that it starts with an open question ("What is 'important' to you about your group work?"); it proceeds intersubjectively; and it is purposely disruptive in ways that provoke discussion and contributions from all members until some resolution is reached. The indicators were generated by working groups ('doers') relating to their work/practice ('doing'), rather than by individual leaders or national or regional representatives in an abstract and intellectual mode.

In this paper the authors present an exploratory study of the indicators produced via this values-based approach compared to the conventional approach used by the UN in order to test their influences to the operationalisation of complex SDG targets. To do this the values-based indicators developed by twenty-one organisations to evaluate their broadly-defined 'ESD' programs are compared to those obtained through the UN process for SDG Target 4.7 which is focused on ESD. The analysis involved first constructing a conceptual framework for Target 4.7 using the method of Hak et al. (2016; 2007), and using it to systematically compare, via coding, the two sets of indicators for their usefulness for operationalising in terms of their potential relevance (item validity), and coverage (sampling validity).

The following sections discuss principles of indicator design, as background to the subsequent description of the conventional approach currently used within the UN SDG process and the values-based approach that we have studied.

4.1.1. Emerging Principles of Indicator Design

The development of indicators for the SDGs should be systematic and informed by theory (Hák et al., 2016; Lu, Nakicenovic, Visbeck, & Stevance, 2015) and should be undertaken in a way that goes beyond 'what *can* be measured using currently available methods' in order to measure 'what *should* be measured' (McCool & Stankey, 2004, pp. 295-296). The strong appeal by McCool and Stankey (2004) that we go beyond the convenient constitutes recognition that the development of indicators is informed by both social and political negotiations (Tamas, in press). Were indicator development purely technical, there would be no risk that the convenient could be substituted for the valid.

Hák et al. (2016) advocate the creation of coherent conceptual frameworks for SDG targets, as a necessary condition for ensuring the validity of linkages between indicators and the facts they are to represent. If the conceptual framework behind SDG targets is not internally coherent, that incoherence will compromise their indicators (see also Tamas, in press). This echoes pre-SDG indicator development guidelines such as those included by Dalal-Clayton and Bass in their resource book *Sustainable Development Strategies* (2002), advocating the creation of a 'framework of parts and aims' as the first essential step in designing an indicator-based assessment. In this context, 'parts' are understood as

118

dimensions, elements or themes, and 'aims' as goals, objectives, principles or criteria. These authors suggest that using a framework of parts and aims as the foundation for indicator development has numerous advantages, e.g., avoidance of measuring the same part twice or omitting essential parts; highlighting parts for which no measurable indicator exists; and showing the relationships between different parts and appropriate weightings for each of them in the development of indices (see also (Grainger, 2012; Gudmundsson, 2003).

Dalal-Clayton and Bass (2002) further recommend the development of a framework of indicators that is *systemic* (organised to illustrate key features of the system and subsystems), *hierarchical* in the sense that the parts are organised into a series of levels, *logical* in the sense of being a series of means and ends (e.g., 'we measure A by measuring B, B by measuring C…'), and readily communicable to non-specialists. More recently, Lozano, Llobet and Tideswell (2016) have emphasised the importance of examining the interlinkages between different indicators within the framework.

Assuming that a given indicator is empirically defensible, it must also be recognised as socially relevant (e.g., policy relevance, relevance to specific stakeholder groups, or applicability at the appropriate level) (Hák et al., 2016). Optimising indicators for both the empirical and the social, in turn, requires clear conceptions of key terms such as 'sustainable', 'efficient' or 'substantial' (Lu et al., 2015)—terms whose ambiguity may have been useful and intentional when the targets were negotiated, but which then pose challenges for operationalisation. These issues can be avoided in narrowly defined SDG areas, but are problematic with complex areas which involve intermeshed concepts, such as ESD.

4.1.2. The Conventional UN Approach to Developing SDG Indicators

The task of deciding what should be measured to evaluate the SDGs at a global level was delegated to the 'Inter-Agency and Expert Group on Sustainable Development Goal Indicators' (IAEG-SDG) by the United Nations Statistical Commission in March 2015. This group consists of representatives from the national statistical offices of 28 'member' countries (consisting of seven each from Africa and Asia, two each from Oceania and the Caribbean, one from North America, three from Central and South America, and five from Europe) (IAEG-SDG, 2016a), as well as observers from non-member countries, regional

commissions, and regional and international agencies (e.g., non-governmental organisations) who can provide technical advice and support (IAEG-SDG, 2016b).

In its March 2016 report, demonstrating the extent to which indicator development is as much socio-political as it is empirical, the IAEG-SDG describes the initial step in the process of indicator development as involving an "open consultation . . . with all countries, regional and international agencies, civil society, academia and the private sector" from August to September 2015 (IAEG-SDG, 2016b) This was followed by an initial coding of all proposed indicators as either 'green' (general agreement) or 'grey' (more discussion required) by the panel. Two limited public consultation processes were then launched—the first, on the 'green' indicators, open for only three days in November 2015 (IAEG-SDG, 2015b), and the second, on the 'grey' indicators, for seven days in December 2015 (IAEG-SDG, 2015a). While the consultations received little or no attention in the mainstream media and academic journals, their output represents a snapshot of the international discourse surrounding SDG indicators at the crucial launch time of the goals. They are noteworthy not only in the sense of who has contributed and what has been said, but also—as we discuss below—because of what and who has *not* been included.

It is apparent from the March 2016 report and related IAEG-SDG documents (IAEG-SDG, 2015a, 2015b, 2016b) that the IAEG-SDG approach to indicator design did not begin with the decomposition of the target into a framework of parts and aims, as advocated by Dalal-Clayton and Bass (2002). Rather, individuals and organisations were invited to comment on draft indicator proposals issued by the IAEG-SDG, and submit their recommendations for alternatives. This raises the possibility that proposed indicators might have been informed as much by participants' cultural backgrounds, institutional affiliations and interests as by the functional need. Insofar as there is no opportunity for critical cross-examination or acknowledgement of the full range of concepts to be covered, the entire process invites contributors to formalise and uncritically advance their own perhaps entrenched ways of thinking about what constitutes 'SD'.

4.1.2.1. Operationalizing 'ESD' through the Conventional Approach

The strategic importance of education as a main Sustainable Development Goal (Goal 4) has been widely acknowledged, and its two-way linkages with other goals outlined, in numerous United Nations flagship reports—to the extent that the goals can be implicitly organised as a conceptual map with education at the centre (Vladimirova & Le Blanc, 2016).

The language of Goal 4 responds to criticisms levelled at the earlier Millennium Development Goals concerned with education, which focused on net primary school enrolment and gender parity and failed to mention quality or to recognise that education continues beyond formal schooling [4]. While the current goal is thought to be an improvement, the term 'quality' is an example of the ambiguous wording critiqued by Lu and colleagues (2015).

In this paper we focus specifically on operationalising Target 4.7, the target which relates most explicitly to education for sustainable development (ESD):

SDG Target 4.7:

"*By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development.*"

We have chosen to focus on this SDG target because its great breadth provides an excellent overlap of subject domain with that available from the values-based approach that we wish to consider.

In examining Target 4.7 we do not adopt a position on what is meant by either 'sustainable development' or 'education for sustainable development' and we fully acknowledge that the term 'ESD' is highly controversial and that its usefulness has been contested, especially by environmental educators (Jickling & Wals, 2008). Nevertheless, we accept that the language of ESD has become so entrenched—not only in the SDGs themselves but also in a vast body of United Nations literature, particularly around the UNESCO 2004–2014 'Decade of Education for Sustainable Development'—that it would be unrealistic to

imagine that current objections alone could secure its replacement with a less politically loaded term (such as 'education for sustainability'). We expand further on these points, with relevant supporting literature, in Appendix A.

While some targets for other SDGs also have relevance to ESD, especially Target 12.8 "ensure that people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature") and Target 13.3 ("improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning"), these are very much narrower in focus than Target 4.7. The complexity of Target 4.7 not only provides better domain overlap for our comparison (see Section 1.3 below), but also epitomises both the challenges of, and the crucial need for, a systematic and theoretically-grounded approach to conceptualisation and operationalisation in the development of complex SDG indicators.

The initial attempt by the IAEG-SDG [44] to operationalise this target, using one indicator, took a very narrow interpretation that focused specifically on scientific knowledge:

Candidate Indicator (UN-IAEG) for Target 4.7:

"Percentage of 15-year old students enrolled in secondary school demonstrating at least a fixed level of knowledge across a selection of topics in environmental science and geoscience. The exact choice/range of topic will depend on the survey or assessment in which the indicator is collected. (Dis-aggregations: sex and location (and others where data are available)."

This indicator is problematic not only because of the narrow scope, but also its promotion of a transmissive rather than a transformative view of ESD (c.f. Jickling & Wals, 2008). The recognition by the IAEG members that this indicator does not fully reflect the concepts found in the target (i.e., lacks sampling validity) is evident in their decision to call for proposals for alternative indicators (IAEG-SDG, 2015a).

A total of 114 separate responses to the public consultation were received for Target 4.7, of which 83 came from civil society organisations (CSOs), mainly non-governmental organisations; 16 from United Nations agencies or national UN-related organisations; nine from national statistical offices; five from universities; and one from a working group

specifically convened for the purpose of reflecting on SDG indicators (the 'Adolescent Girl and SDG Indicators Working Group', which was comprised of representatives from the UN Foundation and six CSOs). Several were replicated identically, but the 71 unique indicators form one set for our comparison and we label them the 'IAEG-SDG' responses set, produced through the conventional, UN, approach. Below we analyse their relevance and coverage for Target 4.7.

4.1.3. Values-Based Approaches to Indicator Design

The second approach and indicator set that we will analyse originated with the ESDinds project (ESDinds: Development of Values-Based Indicators and Assessment Tools for Civil Society Organisations Promoting Education for Sustainable Development), a project funded by the European Commission's Seventh Framework Programme from 2008 to 2011 under the specialist funding scheme 'Research for the Benefit of Specific Groups: Civil Society Organisations' (BSG-CSO) [49]. The two-year project was constructed as a consortium of two academic partners (university research groups) and four civil society organisations promoting ESD in non-formal contexts:

(a) the Alliance of Religions and Conservation, a secular organisation working with 11 major faiths;

(b) the Earth Charter Initiative, a global network of organisations promoting the principles of the Earth Charter, with its secretariat in Costa Rica and strong representation from the global South;

(c) EBBF (ethical business building the future), a global learning community initially registered as a non-governmental organisation in France;

(d) People's Theater (sic), a small German organisation.

The consortium evolved from earlier collaborations and partnerships, and consisted of organisations that viewed their ESD provision as broadly 'values-based'—defining 'values' in the sense of "principles or standards of behaviour", i.e., ethical, spiritual or moral values, as well as "[people's judgement of what is important in life" (Oxford English Dictionary, 2013). The overarching aim was to develop project-level indicators and tools that would not only capture the values and priorities of the participating civil society organisations (CSOs) in their work, but also help them to identify, evaluate and communicate less tangible ESD dimensions.

While the ESDinds process did not set out to produce indicators that covered Target 4.7, we show that its execution within organisations with express interest in ESD was sufficient to do this (see Table 4-1).

Table 4-1. A summary of factors in the ESDinds and IAEG-SDG design approaches which indicate domain differences or similarities for consideration in the comparison analysis

	Values-Based Approach (ESDinds) (N = 125)	UN-IAEG Conventional Approach (N = 114)
Who was asked?	Members of 6 groups, civil society organisations (CSOs): their answers were used to construct the set. Members of a further 15 CSOs: their answers were used to reduce the set by clustering overlapping responses	All countries; regional and international agencies, civil societies, academia and the private sector were asked.
Who answered	As above	114 groups answered, of which 83 CSOs, 16 UN Agencies or related, nine national statistics offices, five universities and one specific working group.
Question responded to	"What is worthwhile, valuable and meaningful to you about your work?"	"Please examine the Candidate Indicator given for SDG Target 4.7 and propose any alternatives you think might be more appropriate."
Boundaries which are contextually implied:	The question was with respect to	The question was with respect to
	a A local group	a Mostly national or international groups
	b Non-formal education	b Formal or non-formal—not specified
	c Assessment of ESD of a collective	c Assessment of ESD of individuals
	d Local SD focus	d Local or global—not specified
	e The present situation and near future	e Present, near or far future—not specified
	f Practice (as opposed to knowledge)	f Practice or knowledge—not specified
	g Holistic approach to assessment	g Non-holistic approach to assessment
	h ESD decomposition via local values	h ESD deconstruction via outputs
	i Obtained intersubjectively within the group	i Process within groups unknown but likely to be non-inter-subjectively

The methodology of the ESDinds project is described fully in the final report (ESDinds, 2011) and by Podger et al. (2010) and is usefully understood to be an example of 'Research through Design' (see Chapter 1 and Chapter 5). In brief, the initial process of developing draft indicators used a multiple case study approach, incorporating diverse methods of data collection—key informant semi-structured interviews, informal interviews, participant observation, and document collection—to explore what group participants found valuable, meaningful and worthwhile in the context of their values-based ESD projects. Six organisations contributed in this way, all affiliated to the four partner CSOs. Thematic content analysis of these large data sets was based around a codebook derived from values and indicators literature, and involved tests for inter-rater reliability and member checking within

all participating CSOs. During the analysis, the wording of each draft indicator was negotiated so that participating organisations found them to be comprehensible, measurable, relevant and locally valid, while attempting to achieve a level of abstraction that would make them generalisable to other contexts. This was in accordance with the 'bottom-up and top-down' approach presented as an example of good practice in indicator development by Reed, Fraser and colleagues (Fraser et al., 2006; Reed & Dougill, 2002; Reed, Dougill, & Baker, 2008; Reed et al., 2006; Reed et al., 2005).

The draft indicators from the different organisations were then clustered intersubjectively to produce a set of 177. This was later reduced to a final reference set of 125 applicable to multiple contexts, including but not limited to non-formal ESD (ESDinds, 2011), via field trials in 15 further organisations, and that set is centred on ESD practice. We thus analyse it for relevance and coverage of Target 4.7, denoting it the ESDinds set, and the approach as 'values-based'. This process produced not rigid indicators, but 'proto-indicators'—reflecting the concept of a 'prototype' in design literature, to refer to concise statements that can serve as templates or triggers for the local development of immediately relevant measurable indicators [33].

Proto-indicator sets have since been developed in a simplified manner for other contexts, including secondary schools (Chapter 5; see also Dahl et al., 2014a, 2014b; Dahl et al., 2014c), universities (Burford et al., 2014; Ribeiro et al., 2016) and community-university research partnerships (Hoover et al., 2015), All the sets have significant overlap, and their indicators reflect the values-based approach. However, none of those have yet achieved the same level of validation through field testing as the ESDinds set.

4.1.4. The Overlap of the Two Domains

The aim of this study is to compare indicators produced from the two approaches to explore their intrinsic differences with respect to useful operationalisation of Target 4.7. Ideally, the two would be developed in parallel, but without time and funds to do this we have made use of what is available, and noted areas where comparisons would not be appropriate or need extra caution. Table 4-1 summarises these, and is referred to in the subsequent design of analysis. The aim is to identify aspects of the indicator sets or the processes by which they are generated which suggest that further research would be useful.

4.2. Background from education literature[15]

4.2.1. Usage and Critiques of 'Education for Sustainable Development' (ESD)

Two decades ago, there were already over 300 definitions of 'sustainable development' and 'sustainability' (Dobson, 1996); now the number may be in thousands. An important manifestation of this ambiguity is the deep division within the environmental education community about the appropriateness, or otherwise, of the term 'education for sustainable development'. Although the UNESCO definition of ESD refers to an education that "empowers learners to take informed decisions and responsible actions for environmental integrity, economic viability and a just society, for present and future generations, while respecting cultural diversity" (UNESCO, 2016), the word 'development' remains inherently problematic for many. It evokes long histories of paternalistic, centrally organised, and often environmentally and socially destructive economic policies (Robinson, 2004). The term 'education for sustainable development' is often felt to imply an endorsement of neoliberal economic growth agendas and the corporate globalisation of education, and a corresponding negation of non-growth-based ideologies such as Arne Naess's concept of 'deep ecology' (Naess, 1973). Indeed, the United Nations resolution which launched UNESCO's 'Decade of Education for Sustainable Development' has been criticised for lacking any explicit reference to ecology or the environment (Jickling & Wals, 2008).

Yet however academics might problematize ESD and call for its replacement with less loaded terms such as 'education as sustainability' or 'sustainable education' (Sterling, 2001), or even 'education consistent with Agenda 21' (Smyth, 1999), there are both pragmatic and ideological reasons for keeping ESD at the forefront of discussions about SDG indicator development. On the pragmatic side, the success of the UN Decade of ESD has resulted in the concept becoming firmly embedded in global discourses: not only in the SDGs themselves, but also in the Muscat Agreement adopted at the 2014 Education for All conference, and

---

[15] This section was included in the published version as 'Appendix A', to comply with journal word limits for the main text. I have reintegrated it into the main text here in order to improve flow and comprehensibility.

intergovernmental declarations on climate change, biodiversity, sustainable consumption and production, and many other issues (UNESCO, 2014). The question of its measurement thus has enormous practical significance.

On the ideological side, any attempt to create a single, universally applicable indicator to represent the entire domain of ESD can be viewed as a way of reinforcing global hegemonies and silencing debate, and as "fundamentally 'mis-educative' and anti-democratic"; and by remaining silent on this issue, we effectively become complicit in it. This is particularly true when the proposed indicator frames ESD in terms of disseminating scientific information (transmissive education), rather than empowering learners as critical, creative and proactive change agents (transformative education) (c.f. Jickling & Wals, 2008)[16].

4.2.2. Review of Literature on (E)SD Competencies

The term 'competency' has been defined by Rychen and Salganik (2001, p. 51) as "the ability to successfully meet complex demands in a particular context through mobilisation of psychosocial prerequisites (including both cognitive and non-cognitive aspects)", and as "a complex action system encompassing cognitive skills, attitudes, and other non-cognitive components". We have chosen to adopt a competencies approach because, as illustrated by the examples below, this framing has become widespread throughout academic and professional literature on 'skills for promoting SD'. However, there are some important caveats associated with the use of this term, owing to its historical roots in "competencies-based education"—which was often narrowly defined as education that sought to prepare people for particular jobs or functions—and its association with discourses of quality and accountability (O'Donoghue & Chapman, 2010, p. 85).

One such caveat is that in listing specific competencies that students might acquire, we must not lose sight of the interconnections between them, nor fall into the trap of conceptualising learning as segments of knowledge that can be ordered in a hierarchical sequence. Another is that focusing on competencies may imply that knowledge is static and

---

[16] What I did not say explicitly in the published paper is that this theoretical argument strengthens the case for democratisation and pluralism in indicator design: see Section 4.4.8.

can only be 'acquired', rather than being something that learners themselves can produce as autonomous thinkers (O'Donoghue & Chapman, 2010, pp. 94-95). A third important point is that competencies can only be measured indirectly, as they are manifested in very specific contexts of behaviour and action (Rychen & Salganik, 2001; Soucek, 1993).

A seminal report to the United Nations Educational, Scientific and Cultural Organisation (UNESCO) by the International Commission on Education for the Twenty-First Century, chaired by the former European Commission President Jacques Delors, distinguishes four 'pillars' for education in the twenty-first century. These are *learning to know* (e.g., learning how to learn, developing critical thinking, acquiring tools for understanding the world, and understanding sustainability concepts and issues); *learning to do* (e.g., acquiring technical and professional training and applying learned knowledge in daily life), *learning to be* (e.g., seeing oneself as the main actor in defining positive future outcomes, developing self-identity and self-knowledge, acting with autonomy and personal responsibility), and *learning to live together* (e.g., understanding other people's values and traditions, cooperating with people, celebrating diversity and coping with conflict) (Delors et al., 1996).

The 'Delors Report' has had a substantial impact on education policy and practice worldwide, and constituted a major theme in more than 50 global conferences (Tawil & Cougoureux, 2013). These four pillars, which have been respectively correlated to 'domain competencies' (or 'subject competencies'), 'methodological competencies', 'personal competencies' and 'social competencies' (Erpenbeck & von Rosenstiel, 2003; Sleurs, 2008), have been applied specifically to ESD as the foundation of the United Nations Economic Commission for Europe (UNECE) report *Learning for the Future: Competences in Education for Sustainable Development* (UNECE, 2011). In this report, these domain, methodological, personal and social competencies are applied across three 'fields', namely 'taking a holistic approach', 'envisioning change' and 'achieving transformation', and respectively framed for ESD educators rather than students as follows: 'The educator understands…' [Learning to Know], 'The educator is able to… ' [Learning to Do], 'The educator is someone who… ' [Learning to Be], and 'The educator works with others in ways that…' [Learning to Live Together]. This report incorporates much more detail than the original UNESCO framework of competencies. To these four main competencies a fifth, *learning to transform oneself and society* (expanded as minimizing ecological footprints, integrating sustainable lifestyles,

creating gender-neutral and non-discriminatory societies, and respecting the Earth and life in all its diversity) was added on the advice of Latin American educators and the United Nations Children's Emergency Fund (UNICEF) and later officially adopted by UNESCO (Combes, 2005).

The Delors-inspired *Learning for the Future* report is cited in the 2012 report of the 2005–2014 Decade on Education for Sustainable Development, *Shaping the Education of Tomorrow*, which notes that a key change during the Decade was the growing recognition not only of the contested nature of ESD, but of "the importance of the capacities, skills, competencies and qualities people need to contribute to transitioning towards a more sustainable world" (UNESCO, 2012, p. 22). The same report also cites a 2011 UNECE evaluation (UNECE, 2011, p. 23) which highlights the "need for distinction between SD competence (e.g., citizen's capacities to contribute to sustainable living both professionally and personally) and ESD competence (e.g., an educator's capacity to help people develop SD competence through a range of innovative teaching and learning practices)".

The Comenius-2 funded CSCT project - 'Curriculum, Sustainable Development, Competences, Teacher Training' - identifies three overall competencies for sustainable development, namely 'teaching/communicating', 'reflecting/visioning' and 'networking', overlaid on five domains of professional competencies (respectively values and ethics, action, knowledge, systems-thinking and emotions) which mirror the Delors 'pillars' to some extent, and can be applied to teachers as individuals, within their institutions, and in wider society (Sleurs, 2008).

Wiek et al. (2011, p. 212) have conducted a systematic review of both academic and grey literature on sustainability competencies, in which they critique the existing literature for its lack of rigor—particularly because it tends to refer to competencies in list form, rather than grounding them in a theoretical understanding of sustainability and highlighting interlinkages in the form of a conceptual framework. Their systematic review of 28 academic sources (journal papers and books) and 15 grey literature sources (reports and White Papers) resulted in the identification of five main clusters of key competencies for sustainability:

- systems thinking competence, the ability to analyse the dynamics of complex social-ecological systems;

- anticipatory competence, i.e., the ability to create, analyse and evaluate what they term 'rich pictures' of the future;

- normative competence, also referred to as value(s)-focused thinking, which focuses on "the ability to collectively map, specify, apply, reconcile, and negotiate sustainability values, principles, goals, and targets" (p. 10);

- strategic competence, which ensures that learning is translated into effective policies, programs and action plans; and

- interpersonal competence, which refers to the ability to motivate, enable and facilitate participatory sustainability research and collaborative problem-solving, to celebrate diversity, and critically evaluate different positions and perspectives. This is regarded as a cross-cutting issue, as the involvement of multiple actors in sustainability problems makes it essential for collective strategies to be employed in trying to solve them.

Wiek et al. argue that these five categories, which they term *key competencies*, are essential for sustainability and should therefore be distinguished from 'basic' competencies such as critical thinking or communication skills in a more generic sense.

More recent work largely reiterates similar points, rather than making substantive new contributions to the definition of ESD competencies. Murga-Menoyo (2014), for example, echoes Wiek et al. (without citing them directly) in differentiating 'generic traditional competencies' (such as problem-solving, critical and self-critical skills, or the ability to work in a team) from 'generic ompetencies for sustainability'. Murga-Menoyo states that generic traditional competencies, which they characterise as being either instrumental, interpersonal or systemic, are a prerequisite for the achievement of the latter.

Table 4-2, below, correlates post-2012 references on ESD competencies (Cebrián & Junyent, 2015; Education for All, 2014; Murga-Menoyo, 2014) with the framework developed by Wiek et al. (2011). Within the domain of critical thinking, Cebrián and Junyent specifically highlight the importance of reflecting critically on the values, beliefs and worldviews underpinning current ways of knowing and understanding, and co-constructing new shared meanings (Sterling, 2001; A. E. Wals & Corcoran, 2006).

Table 4-2. Sustainability competencies identified by Wiek et al. and more recent authors

Category from Wiek et al. (2012)	Related Framings from Recent Literature	References
Systems-thinking competencies	Critical contextualisation of knowledge (interrelating social, economic and environmental issues, local and global levels)	Murga-Menoyo [102]
	Problem solving, critical thinking	Cebrian and Junyent ([103], p. 2769) Education for All (EFA) Global Monitoring Report 2013–2014 ([104], p. 295)
	Working and living with complexity	Cebrian and Junyent [103]
Strategic competencies	Sustainable use of natural resources and prevention of negative social and environmental impacts	Murga-Menoyo [102]
	Action competence: decision-making, participation and action for change	Cebrian and Junyent ([103], p. 2769)
Interpersonal competencies	Communication, co-operation […], conflict resolution, leadership and advocacy	Education for All (EFA) Global Monitoring Report 2013–2014 ([104], p. 295)
	Participation in community processes towards sustainability	Murga-Menoyo [102]
	Managing emotions and concerns	Cebrian and Junyent [103]
	Establishing interdisciplinary dialogues	Cebrian and Junyent [103]
Normative competencies	Values clarification	Cebrian and Junyent [103]
	Enacting sustainability values and applying ethical principles in both professional and personal contexts	Murga-Menoyo [102]
Anticipatory competencies	Visioning futures or alternative scenarios	Cebrian and Junyent [103]

For conceptual simplicity and ease of communication, the Wiek model could be entirely contained within that of UNESCO/UNECE. Systems-thinking competencies, for example, might be seen as a subset of 'learning to know'; anticipatory competencies as a subset of 'learning to transform oneself and society'; and normative competencies as a subset of 'learning to be'. The Wiek categories of interpersonal and strategic competencies can respectively be viewed as synonymous with 'learning to live together' and 'learning to do'. We would argue, however, that the political importance of the largely neglected (sub-)categories of systems-thinking, anticipatory and normative competencies negates any pragmatic advantage gained by subsuming them within larger categories. Operationalisation of systems-thinking, anticipatory and normative competencies could have wide-reaching implications, especially in terms of conceptual use of any resulting indicators—inspiring people to change the ways that they think, plan and reflect on values.

4.3. Methods

In this section we briefly outline the approach used to develop a conceptual framework for Target 4.7 and then describe the analysis conducted to compare the IAEG-SDG and ESDinds indicators, from the conventional and values-based approaches respectively, in terms of their usefulness to the operationalisation of this target. Reference is made to their relevance and coverage of different components which would impact on the overall item validity and sampling validity of any final indicator(s) developed.

4.2.1. Creating a Conceptual Framework for Target 4.7

The process of creating a conceptual framework for Target 4.7 proceeded in three stages: (1) defining two sub-targets; (2) subdividing the skills sub-target, by identifying broad categories of sustainability competencies derived from academic and professional literature; and (3) identifying smaller clusters of skill competencies within those categories where applicable, again with direct reference to literature. A broad overview is given below and further details in Appendix B.

4.2.1.1. Top Level: Sub-Targets

Hak et al. (2016, p. 570) provide an example of the process of defining sub-targets by breaking down Target 12.3, "By 2030 achieve sustainable management and efficient use of natural resources", into two sub-targets, "sustainable management of natural resources" and "efficient use of natural resources". Each of these, in turn, can be subdivided, e.g., efficient use of raw materials (comprising fuels, minerals, metals and biomass), efficient use of food, efficient use of water, and efficient use of land. With each recursively created level, the sum of all the sub-targets constitutes the super-ordinate.

The application of the same approach to Target 4.7 is frustrated by the lack of natural points where it may be cleanly separated into mutually exclusive and collectively exhaustive sub-targets. In principle, the target could be split cleanly into two overall dimensions relating to learning outcomes—*all learners acquire the knowledge needed to promote sustainable development*, and *all learners acquire the skills needed to promote sustainable development*.

132

[However, this is problematic for several reasons, which are described in the Appendix to Chapter 4. Accordingly, we have proposed the following wording for the sub-targets, whose derivation is explained in the Appendix:][17]

> *Sub-Target 1: All learners and educators involved in organised learning activities, both formal and non-formal, acquire knowledge needed to promote sustainable development.*
>
> *Sub-Target 2: All learners and educators involved in organised learning activities, both formal and non-formal, acquire skills needed to promote sustainable development.*

A case could also be made for including a third or alternative sub-target that can be operationalised as multiple process indicators, e.g., "All learners enrolled in formal and non-formal organised learning activities receive education for sustainable development and sustainable lifestyles" (or human rights, gender equality, peace and non-violence, etc.), on the grounds that in certain countries it may be logistically difficult or impossible to assess learning outcomes for ESD. Where this is the case, the only way of collecting meaningful data would be to focus on the provision of relevant learning activities. In this paper, however, we focus on operationalising the target as it is currently worded, with its focus on learning outcomes, only returning to this alternative sub-target in the discussion.

4.2.1.2. Second Level Categories: Competencies

We realised there was no need to conceptualise in great detail 'knowledge that enables people to promote sustainable development' since actions are local and contexts vary, resulting in the knowledge required to promote 'SD' (however it might be defined) being idiosyncratic to each locality. Attempting to specifically catalogue this knowledge for a global level of application would be counter-productive, hence the need for localisable assessments such as the 'Sustainability Literacy Test' (SuLiTest) developed by UNEP and other agencies. On the other hand, this same SuLiTest is helpful as an initial proxy measure for SD-related knowledge at a general level, and we advocate its use as a starting point even though it has

---

[17] These two sentences were accidentally omitted from the published version of the paper and have been reinstated here.

significant limitations, focuses on a minimum level of basic knowledge, and requires further required to develop higher-level assessments (Carteron & Decamps, 2014).

In contrast, we are able to develop great detail in the conceptualisation of skills for Sub Target 2. Despite being largely overlooked by the IAEG-SDG in their indicator development process, the skills dimension features prominently in ESD in both professional and academic literature on ESD—usually framed in terms of 'sustainability competencies'. We thus expanded our framework in the skills dimension through a sub-study of significant works relating to skills that enable people to promote sustainable development. Ultimately we produced a hybrid framework which covered the key aspects of the competencies discussed elsewhere. It was comprised of two accepted conceptual frameworks, one from professional literature (Delors et al., 1996; United Nations Economic Commission for Europe (UNECE), 2011) and the other from academic literature (Wiek, Withycombe, & Redman, 2011). Both are framed in terms of 'sustainability competencies', also referred to as 'ESD competencies', although we note that the term 'competency' as defined by Rychen and Salganik (2001, p. 51) includes not only skills but also other cognitive and non-cognitive components, such as attitudes). Appendix A sets out the justification for, and caveats associated with, the use of the competencies lens (and of these specific references) within a broader literature review on ESD, while Appendix B provides further detail on the actual construction of our conceptual framework. The key concepts of the two frameworks used are summarised below:

> **UNESCO/UNECE model**: A seminal report to the United Nations Educational, Scientific and Cultural Organisation (UNESCO) by the International Commission on Education for the Twenty-First Century, chaired by the former European Commission President Jacques Delors (1996), distinguishes four 'pillars' for education in the twenty-first century. To these four pillars, which have been applied specifically to ESD as the foundation of the United Nations Economic Commission for Europe (UNECE) report *Learning for the Future: Competencies in Education for Sustainable Development* (2011), a fifth was later added by UNESCO (see Appendix A), resulting in the following set of competencies:

- *learning to know* (e.g., learning to learn, developing critical thinking, acquiring tools for understanding the world, and understanding sustainability concepts and issues);

- *learning to do* (e.g., acquiring technical and professional training and applying learned knowledge in daily life);

- *learning to be* (e.g., seeing oneself as the main actor in defining positive future outcomes, developing self-identity and self-knowledge, acting with autonomy and personal responsibility);

- *learning to live together* (e.g., understanding other people's values and traditions, cooperating with people, celebrating diversity and coping with conflict);

- *learning to transform oneself and society* (e.g., building non-discriminatory societies, reducing ecological footprint, integrating sustainable lifestyles and promoting social solidarity).

Wiek model: Through a systematic literature review, Wiek et al. (2011) identify five broad categories of key competencies for sustainability:

- *systems thinking competence*, the ability to analyse the dynamics of complex social-ecological systems;

- *anticipatory competence*, i.e., the ability to create, analyse and evaluate what they term 'rich pictures' of the future;

- *normative competence*, also referred to as value(s)-focused thinking, which focuses on "the ability to collectively map, specify, apply, reconcile, and negotiate sustainability values, principles, goals, and targets" (ibid, p. 10);

- *strategic competence*, which ensures that learning is translated into effective policies, programs and action plans; and

- *interpersonal competence*, which refers to the ability to motivate, enable and facilitate participatory sustainability research and collaborative problem-solving, to celebrate diversity, and critically evaluate different positions and perspectives. This is regarded as a cross-cutting issue, as the involvement of multiple actors in sustainability problems makes it essential for collective strategies to be employed in trying to solve them.

While the literature on competencies in ESD reveals a complex, multifaceted and rather low-consensus picture comprising both affective and cognitive elements (see Appendix A), we find that most of this complexity is covered within the two models described by UNESCO/UNECE and Wiek et al., respectively. For the practical purpose of this paper, which requires a framework adequate to roughly compare potential relevance (item validity) and coverage (sampling validity) of indicators, we have combined the main concepts of these two models into a single eight-category framework of ESD competencies, described in detail in Appendix B and illustrated in Figure 4-1, (1)–(8) below.

4.2.1.3. Third Level: Specific Clusters of Competencies

Of the eight broad competency categories, five contained further distinctive sub-dimensions, as shown in the bottom half of Figure 4-1, closely following the text of a UNECE infographic [68]. In order to balance rigour (in the sense of including a sufficient number of codes to provide a valid and reliable recognition of all analytically relevant data) with usability, we worked at the level of clusters. Our conceptual framework for Target 4.7 thus consisted of a Sub-Target for Knowledge with only one level of division of concepts, and a Sub-Target for Skills with three levels of division of concepts (which are shown in Figure 4-1).

4.2.2. Exploring the Usefulness of the Operationalisation of Target 4.7 through Two Contrasting Approaches

In considering how to compare the outputs from the two approaches for generating indicators, and to identify important aspects which can feed into future research design, attention must be paid to the incomplete overlap of approach domains. Had the aim of this work been to evaluate two approaches across equivalent domains, we could have done tests of relevance (item validity) and coverage (sampling validity) of the two sets of indicators using statistics. However, the authors realised it was more useful to obtain a detailed and rich overview of interesting aspects of not only the indicators but also the processes which produced them. For example, although we have described the non-conventional approach taken by ESDinds as 'values-based', it also has clear aspects of intersubjectivity, holism and local-level operationalisation, which might prove more important than the values lens itself. Similarly, the IAEG-SDG process could be described as involving individual representatives of

organisations, rather than working groups of members; direct copying of indicator proposals by other respondents; and a stronger motivation for the participating organisations to push political agendas, such as sexuality or HIV/AIDS education, rather than focusing on the specific Target 4.7.

Figure 4-1. Original conceptual framework for Target 4.7 of the Sustainable Development Goals, created by combining models developed by Wiek et al. (2011) (competency domains shown in grey), and United Nations Educational, Scientific and Cultural Organisation (UNESCO)/United Nations Economic Commission for Europe (UNECE) (2011) (competency domains shown in colour with the respective sub-domains below them).

Target 4.7 of the Sustainable Development Goals:

"By 2030, ensure that *all learners acquire the knowledge and skills needed to promote sustainable development (SD),including, among others, through educatio for SD and sustainable lifestyles, human rights, gender equality, promotio of a culture of peace and non-violence, global citiz nship and appreciation of cultural diversity and of culture's contributio to SD.*"

SUB-TARGET 2: All learners and educators involved in organised learning activ tie, b oth formal and non-formal, acquire **skills** needed to promote SD

(1) Strategic competencies: "collectively design and implement interventions, transitions, and...strategies"	(2) Systems-thinking competencies (SYS): "collectively analyse, complex systems across different domains (society, environment, etc) and different scales	(3) Normative competencies (NOR): "collectively map, specify, apply, reconcile and negotiate sustainability values, principles, goals &	(4) Anticipatory Competencies (ANT): "collectively analyse, evaluate and craft rich 'pictures' of the future related to sustainability issues and sustainability problem-solving	(5) Interpersonal competencies: "motivate, enable and facilitate collaborativ and participatory sustainability research and problem solving"
[equivalent to] **Learning to Do (L2D)**	**(6) Learning to Know (L2K)**	**(7) Learning to Be (L2B)**	**(8) Learning to Transform Self and Society (L2T)**	*[equivalent to]* **Learning to Live Together (L2L)**

L2D_ACT Be an actor as well as a thinker

L2D_APP: Apply learned knowledge in daily life

L2D_CRE Act creatively and responsibly in one's environment

L2D_ISS Understand and act on global and local sustainable development issues

L2D_TRA Acquire technical and professional training

L2K_CON Understand sustainability concepts and issues

L2K_CRI Critical thinking

L2K_LEA Learn to learn

L2K_TAS Acquire a taste for learning throughout life

L2K_TOO Acquire tools for understanding the world

L2B_AUT Be able to act with greater autonomy, judgment and personal responsibility

L2B_DEF See oneself as the main actor in defining positive outcomes for the future

L2B_EXP Encourage discovery and experimentatio

L2B_SEL Develop one's personality, self-knowledge, self-identit and self-fulfilment

L2B_VAL Acquire common global values

L2T_DEM Promote democracy in a society where peace prevails SDGs: n=2 ESDinds: n=2

L2T_FOO Promote behaviors and practices that minimise our ecological footprint on the world around us

L2T_GEN Work towards a gender-neutral, non-discriminatory society

L2T_LIF Develop the ability and will to integrate sustainable lifestyles for ourselves and others

L2T_RES Be respectful of the Earth and life in all its diversity

L2T_SOL Act to achieve social solidarity

L2L_CNF Be able to cope with situations of tension, conflict, violence, exclusion and terrorism

L2L_DIV Tolerate, respect, welcome, embrace and even celebrate difference and diversity in people

L2L_ECN Respond constructively to the cultural diversity and economic disparity found around the world

L2L_PAR Participate and co-operate with others in increasingly pluralistic and multi-cultural societies

L2L_UND Develop an understanding of other people and their histories, traditions, beliefs, values and cultures

Another consideration was whether to focus on 'ideal single indicators' which could potentially cover the entire remit of Target 4.7, or the combined contribution of the output set of indicators from each approach in terms of its 'helpfulness towards operationalisation'. Since the two approaches favoured these differently, both were taken forward so as not to bias either: both provide pathways to producing good coverage.

There were 110 responding organisations in the IAEG-SDG process, providing a total of 114 indicator proposals, but a number of these proposals were repeated multiple times. Our approach required considering only the 71 unique indicators (retaining those with small variations in wording as separate items). The ESDinds set contained 125 proto-indicators which had already been clustered from others, and thus were already unique. Due to the large number of indicators in each set (71 and 125 respectively), and the complexity of the conceptual framework for Target 4.7 (see Figure 4-1), we developed a coding and scoring system to evaluate relevance and coverage for both individual indicators and sets of indicators. We outline this below, using the summary tables to illustrate where less direct considerations must be used, where a lack of domain overlap occurs.

4.2.2.1. Codebook Development and Scoring Methodology of Content Analysis

In order to facilitate the content analysis of both datasets, we developed a codebook from the conceptual framework outlined above. The scoring systems are given in Tables 4-3 and 4-4. By comparing the content of each indicator proposal (IAEG-SDG) or proto-indicator (ESDinds) text with the text in the codebook, an initial subjective judgement of fit was made by the first author, and the coded datasets were subsequently checked by the other authors.

Table 4-3. Scoring method used across the conceptual framework for Sub-Target 1 (Knowledge) of Sustainable Development Goals (SDG) Target 4.7 for relevance and coverage. Note that this Sub-Target is not covered in the values-based approach (ESDinds).

	Relevance	Coverage
Holistic knowledge OR The Sustainability Literacy Test	1 point	2 points
CONTENT x8: 1. Education for Sustainable Development 2. Education for sustainable lifestyles 3. Human rights 4. Gender equality 5. Culture of peace and non-violence 6. Global citizenship 7. Appreciation of cultural diversity 8. Culture's contribution to sustainable development	1 point if AT LEAST ONE content area is assessed	1 point per content area judged to be PARTIALLY covered and 2 points per content area judged to be SYSTEMATICALLY OR RIGOUROUSLY covered
NONE of the above	0 points	0 points

Table 4-4. Scoring method used across the conceptual framework for Sub-Target 2 (Skills) of SDG Target 4.7 for relevance and coverage. The coloured area indicates domains not covered in the values-based approach case- studied in this work (Anticipatory Competencies).

	Relevance	Coverage
	1 point awarded	1 point per competency domain that is PARTIALLY covered, i.e., the indicator covers AT LEAST ONE of its sub-clusters
COMPETENCIES x8:	if ANY of these competency domains are assessed	
L2D Learning to do/Strategic Competencies		2 points per competency domain that is SYSTEMATICALLY OR RIGOROUSLY covered, i.e., the indicator covers ALL of its sub-clusters
L2K Learning to know		
L2B Learning to be		
L2T Learning to transform oneself and society		
L2L Learning to live together/Interpersonal		
Normative Competencies	1 point awarded if ANY of these competencies are assessed (only if not already awarded for UNECE competencies)	1 point per competency domain judged to be PARTIALLY covered *.
Anticipatory Competencies		
Systems Thinking Competencies		2 points per competency domain that the coder judges to be SYSTEMATICALLY OR RIGOROUSLY covered *
NONE of the above	0 points	0 points

* According to the description by Wiek et al.

We used the codebook to analyse the relevance of individual indicators for Sub-Target 1 (Knowledge) and Sub-Target 2 (Skills) respectively, and then to explore the coverage of sub-targets, both at the level of individual indicators and across each dataset as a whole:

- In evaluating **relevance for Sub-Target 1 (Knowledge) at the level of individual indicators**, we scored an indicator as 1 ('relevant') if it explicitly measured the acquisition of knowledge in relation to either (a) at least one of the eight content areas mentioned in the target; or (b) the Sustainability Literacy Test. For this purpose, the terms 'understanding' and 'proficiency' were taken as synonymous with 'knowledge'. Indicators that measured knowledge acquisition only in the limited sense of environmental science and/or geoscience (as included in the original indicator proposal, which the IAEG-SDG had already acknowledged as inadequate for operationalising the target) and those that did not measure knowledge acquisition at all were scored as 0 ('not relevant') for Sub-Target 1. Table 4-3 illustrates and summarises the scoring methods.

- To evaluate **relevance for Sub-Target 2 (Skills) at the level of individual indicators,** 'proficiency' was not taken as synonymous with 'skills', but the word 'skills' was not essential in itself for scoring as 1. Rather, the coder made a judgement, on the basis of the whole indicator text, as to whether the indicator would measure the acquisition of skills in one or more competency domains. Indicators measuring related aspects— such as the creation of appropriate environments for skill acquisition, the existence of policies or structures intended to support ESD skills development, or the percentage of schools providing a certain type of education—were not scored as 1 unless there was clear evidence that the indicator requires the measurement of skill-related learning outcomes for specific individuals or groups in those settings.

- In our evaluation of **coverage of Sub-Targets at the level of individual indicators**, we highlighted the specific content areas (for Sub-Target 1: Knowledge) or competency domains and sub-domains (for Sub-Target 2: Skills) that were hit by each indicator, scoring 1 for 'yes' and 0 for 'no' in each case. (Although there is no particular reason or advantage for a given indicator to score against more than one category, those

scoring highly might be considered more potentially useful as a main indicator for Target 4.7, whereas others might have low coverage but perhaps provide innovation in how to deal with a specific and potentially challenging aspect of ESD.)

- In assessing **coverage of Sub-Targets at the level of the whole dataset**, we assigned an overall category score of 1 point per content area (for Sub-Target 1: Knowledge) or competency domain (for Sub-Target 2: Skills) that we judged to be partially covered, and 2 points per content area or competency domain that we judged to be systematically or rigorously covered. The definitions of partial coverage and systematic/rigorous coverage are provided in Tables 4-3 and 4-4.

One complication occurred in the above process: it was found that one of the Skills sub-domains—namely L2K_CON ('Understand sustainability concepts and issues')—was so close to a description of 'Knowledge' that was not possible to clearly allocate indicators to it. A distinction was implied between knowledge and understanding, but in practice the indicators did not provide such distinctions: for example, it seemed to attract most of the same qualifiers as for Knowledge. We thus excluded this sub-domain completely from analysis.

4.2.2.2. Consideration of Domain Overlaps and Non-Overlaps

A summary of domain differences between the IAEG-SDG and ESDinds indicator development approaches was provided in Table 4-1. The impact of those differences on the analysis for different components of the conceptual framework is illustrated by the shading shown in Tables 4-3 and 4-4 above. Indicators arising from ESDinds were formed through the aggregation of more specific indicators, because part of the process of forming a transferable set across multiple organisations involved this. One consequence is that any natural mention of a specific knowledge area would have been subsumed into any parallel indicator relating to the practice of that knowledge—i.e., a related competency. Thus, with the coding rules as defined above, the ESDinds dataset is not expected to score in Sub-Target 1 (Knowledge) at all, as an artefact of its construction, meaning that the entire Sub-Target cannot be used for comparisons.

Similarly, the ESDinds approach was focussed on the present and near-future for the groups involved, and not the far future, and thus anticipatory competencies are not expected to be covered in its output indicators. Both of these domains could easily be incorporated into future work using an ESDinds-type approach, with very little extra effort. Thus, they cannot be properly considered to be zero scores. However, in this retrospective study and with the specified scoring system, only zero scores are expected for Knowledge. The IAEG-SDG indicators were still processed, to allow consideration of its overall relevance of the approach to both Sub-Targets (Knowledge and Skills).

4.3. Results

In the sections that follow, we present overview summaries of the study findings; compare the item validity and sampling validity of the outputs of the two approaches where their domains overlap (i.e., Skills, other than Anticipatory Competencies); and compare sampling validity across each dataset as a whole at the third level of the conceptual framework, i.e., coverage of competency sub-domains. Specific examples from the coded text of indicators from both conventional and values-based approaches can be found in Appendix C, which readers would benefit from browsing[18]. The full coded datasets for IEAG-SDG and ESDinds are provided as Tables S1 and S2, respectively, in the Supplementary Material[19].

4.3.1. Overview of Relevance (Item Validity)

Table 4-5 provides an overview of the number of useful indicators for the operationalisation of Target 4.7, as produced by the two approaches. It is striking that a significant number of them were not useful for either the Knowledge or the Skills sub-targets.

---

[18] The published paper also included an 'Appendix D', containing three figures which respectively comprised graphical representations of the coverage analysis for Sub-Target 1 for the IAEG-SDG consultation (Figure D1), Sub-Target 2 for the IAEG-SDG consultation (Figure D2), and Sub-Target 2 for the ESDinds proto-indicators (Figure D3). I have omitted this appendix from this thesis because the figures were created by a co-author and are not central to the understanding of the results, described verbally below. Subsequent references to Appendix D, and Figures D1-D3 respectively, have been removed.

[19] The following documents are available online at www.mdpi.com/2071-1050/8/9/861/s1: Table S1: Full coding of responses to IAEG-SDG public consultation on indicators for Target 4.7 of the Sustainable Development Goals; Table S2: Full coding of ESDinds/WeValue indicators.

Given the specific request of the IAEG for responses to focus on Target 4.7, greater relevance might be expected. The second most striking point is that the UN-IAEG approach produced mostly knowledge-based indicators—but many of the respondents had expertise in ESD, and should have been aware of the skills aspects, so it is perhaps surprising that those were underrepresented. Thirdly, the ESDinds approach produced a large number of indicators generally (in the Skills category only—as predicted in Section 2.2.2). Given the original, local question they were designed to address, it is surprising that so many are relevant.

Table 4-5. An overview of the numbers of useful indicators for SDG Target 4.7 produced by the two indicator development approaches (conventional approach exemplified by Inter-Agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDG) public consultation, and values-based approach exemplified by ESDinds).

	IAEG-SDG Public Consultation Responses (Conventional Approach)	ESDinds Proto-Indicators (Values-Based Approach)
	Total N = 71 items	Total N = 125 items
Number of indicators with:		
item validity score of 0 (not useful):	27 (38%)	56 (45%)
item validity score of 1 (useful for only ONE Sub-target):	30	69
Sub-Target 1 (knowledge acquisition only)	22	n/a
Sub-Target 2 (skills acquisition only)	8	69
item validity score of 2 (useful for both sub-targets)	14	n/a

4.3.2. Overview of Coverage (Sampling Validity)

For Sub-Target 1 (Knowledge), the majority of relevant responses from the IAEG-SDG process covered only one or two content areas, with the most common ones being 'gender equality' and 'appreciation of cultural diversity'. It was notable that most of the candidate indicators from the IAEG-SDG process were concerned only with the acquisition of knowledge about sexual and reproductive health. These focused on measuring the outcomes of comprehensive sexuality education through two standard responses, proposed by UN agencies and repeated multiple times by other contributors. Two responses focused only on human rights. Some other responses mentioned knowledge in geoscience and environmental science—content areas mentioned in the target—but lacked specificity about what to assess.

An exceptional contribution was provided by Brookings USA, which included detailed proposals relating to many knowledge domains (and skills, as discussed below). Only one response, from the United Nations Environment Programme (UNEP), called for universal application of the more holistic Sustainability Literacy Test (Carteron & Decamps, 2014) across the tertiary education sector. As anticipated, none of the ESDinds values-based proto-indicators mentioned knowledge acquisition at all. This was an expected artefact of the methods used to aggregate them into a set—see Section 4.2.2.2.

For Sub-Target 2 (Skills), there were far fewer relevant indicators from the IAEG-SDG process, and they generally had poor coverage of Target 4.7. A number of UN consultation responses referred to the concept of "life skills based HIV [human immunodeficiency virus] and comprehensive sexuality education", and some made a brief mention of, for example, vocational skills. Only a minority (n = 22, or 31%) included proposals for indicators that would actually measure the acquisition of skills relevant to content areas outlined in the target. In particular, it is striking that the coverage and number of indicators for the Sub-Target: Skills cluster of "Learning to Know" is extremely low, in contrast to the high coverage found in the Sub-Target: Knowledge. This suggests a narrow view of the nature of knowledge, dominated by content rather than ways of knowing such as critical thinking or learning to be a self-driven learner and thinker. We explore this more fully in the Discussion section below. It is also interesting to note the unusually high number of hits for the sub-domain L2L_DIV within the competency domain of 'learning to live together', i.e., "Tolerate, respect, welcome, embrace, and even celebrate difference and diversity in people.

The results for the same Sub-Target 2 (Skills) from the ESDinds process had a very different profile. As expected, most 'hit' only one competency sub-category, as they were designed to be specific. However, as a set they produce very good coverage of Target 4.7, and in particular in the area that the IAEG-SDG process was weak i.e., in "Learning to Know". Interestingly, like the other approach there is a large number of diversity-related indicators, but unlike it, there are also a large number relating to L2K_ACT (converting knowledge into action) and L2L_PAR (participation). It is also interesting that both approaches did not produce any indicators relating to the Anticipatory Competencies category.

146

The overall coverage score for the full IAEG-SDG dataset was 7 out of a possible 16 points, representing partial coverage of all competency domains other than Anticipatory Competencies. For ESDinds, the overall coverage score was 9, representing systematic/rigorous coverage of Learning to Be and Learning to Transform Oneself and Society and partial coverage of all the other competency domains, with the exception of Anticipatory Competencies. The breakdown of results by sub-domain is shown in Table 4-6.

4.3.3. Sensitivity Analysis

Before proceeding to deeper discussion, we perform an analysis on the data to ensure that it is not inappropriately sensitive to small variations in the process by which each was created. To this end we removed the most significant contributor to each set of indicators. For the IAEG-SDG data it was noted that one particular respondent provided particularly useful and considered indicators—Brookings USA. These indicators achieved high coverage and relevance, and in that respect were anomalous to the other contributions. For a sensitivity test this single, most significant respondent's contributions were removed, and the coverage obtained by the IAEG-SDG approach was re-calculated, as in Table 4-7 below. Similarly, the highest contributing indicator in the ESDinds set was removed.

Table 4-7 shows that the IAEG-SDG score reduced from 7 to 5, where the highest possible score was 16. The ESDinds set performed better, reducing only from 9 to 8.

Table 4-6. Number of indicators coded in each competency domain for candidate indicators from the IAEG-SDG public consultation ('IAEG') and proto-indicators from the ESDinds project ('ESDinds'). Coverage is scored as 0 (none), 1 (partially covered), or 2 (fully or rigorously covered) for each domain.

	Learning to Do					Learning to Know				Learning to Be					Learning to Transform						Learning to Live Together					NOR	ANT	SYS	TOTAL SCORE
	L2D_ACT	L2D_APP	L2D_CRE	L2D_ISS	L2D_TRA	L2K_CRI	L2K_LEA	L2K_TAS	L2K_TOO	L2B_AUT	L2B_DEF	L2B_EXP	L2B_SEL	L2B_VAL	L2T_DEM	L2T_FOO	L2T_GEN	L2T_LIF	L2T_RES	L2T_SOL	L2L_CNF	L2L_DIV	L2L_ECN	L2L_PAR	L2L_UND				
IAEG-SDG (full)	3	2	0	1	4	1	0	0	0	2	2	0	1	2	2	1	3	2	0	1	2	9	0	4	2	4	0	1	
Coverage IAEG-SDG			1				1					1					1						1			1		1	7
ESDinds (full)	#	1	0	0	0	1	2	1	0	7	1	2	9	6	2	8	1	1	4	1	3	#	0	#	2	5	0	1	
Coverage ESDinds			1				1					2					2						1			1		1	9

Table 4-7. Sensitivity analysis for coding of candidate indicators from the IAEG-SDG public consultation ('IAEG') and proto-indicators from the ESDinds project ('ESDinds'). Coverage is scored as 0 (none), 1 (partially covered), or 2 (fully or rigorously covered) for each domain. In each case the most critical item (MCI) has been removed from the analysis.

The figures for the full dataset are also included for the purpose of comparison.

	Learning to Do					Learning to Know				Learning to Be					Learning to Transform						Learning to Live Together								
	L2D_ACT	L2D_APP	L2D_CRE	L2D_ISS	L2D_TRA	L2K_CRI	L2K_LEA	L2K_TAS	L2K_TOO	L2B_AUT	L2B_DEF	L2B_EXP	L2B_SEL	L2B_VAL	L2T_DEM	L2T_FOO	L2T_GEN	L2T_LIF	L2T_RES	L2T_SOL	L2L_CNF	L2L_DIV	L2L_BCN	L2L_PAR	L2L_UND	NOR	ANT	SYS	TOTAL SCORE
IAEG-SDG (full)	3	2	0	1	4	1	0	0	0	2	2	0	1	2	2	1	3	2	0	1	2	9	0	4	2	4	0	1	
Coverage IAEG-SDG			1				1					1					1					1				1	0	1	7
IAEG-SDG (MCI removed)	2	1	0	0	4	0	0	0	0	1	1	0	0	2	2	1	2	2	0	0	1	6	0	2	0	3	0	0	
Coverage IEAG-MCI			1					0				1					1					1				1	0	0	5
ESDinds (full)	#	1	1	0	0	0	2	1	0	7	1	2	9	6	2	8	1	1	4	1	3	#	1	#	2	5	0	1	
Coverage ESDinds			1					1				2						2					1			1	0	1	9
ESDinds (MCI removed)	#	1	1	0	0	0	2	1	0	7	1	2	9	6	2	8	1	1	4	1	3	#	0	#	2	5	0	0	
Coverage ESDinds-MCI			1					1				2						2					1			1	0	0	8

149

4.4. Discussion

4.4.1. Comparisons across Approaches

Figure 4-2 facilitates a broader comparison of the contribution of the two approaches to the coverage of Target 4.7 in domain where they overlap—Skills. This in turn raises interesting questions about the intrinsic nature of the approaches and their potential use in future sustainability processes. Of particular interest is the low number of IAEG-SDG indicators in Learning to Know; the high absolute number of ESDinds indicators generally and in particular in Learning to Do and Learning to Be; and the fact that the highest relative number for both approaches is in Learning to Live Together.

Figure 4-2. A comparison of the contribution of the two approaches to the coverage of Target 4.7 in the most appropriate domain—Skills. It also illustrates the different competency profiles of the two approaches, which deserve reflection.
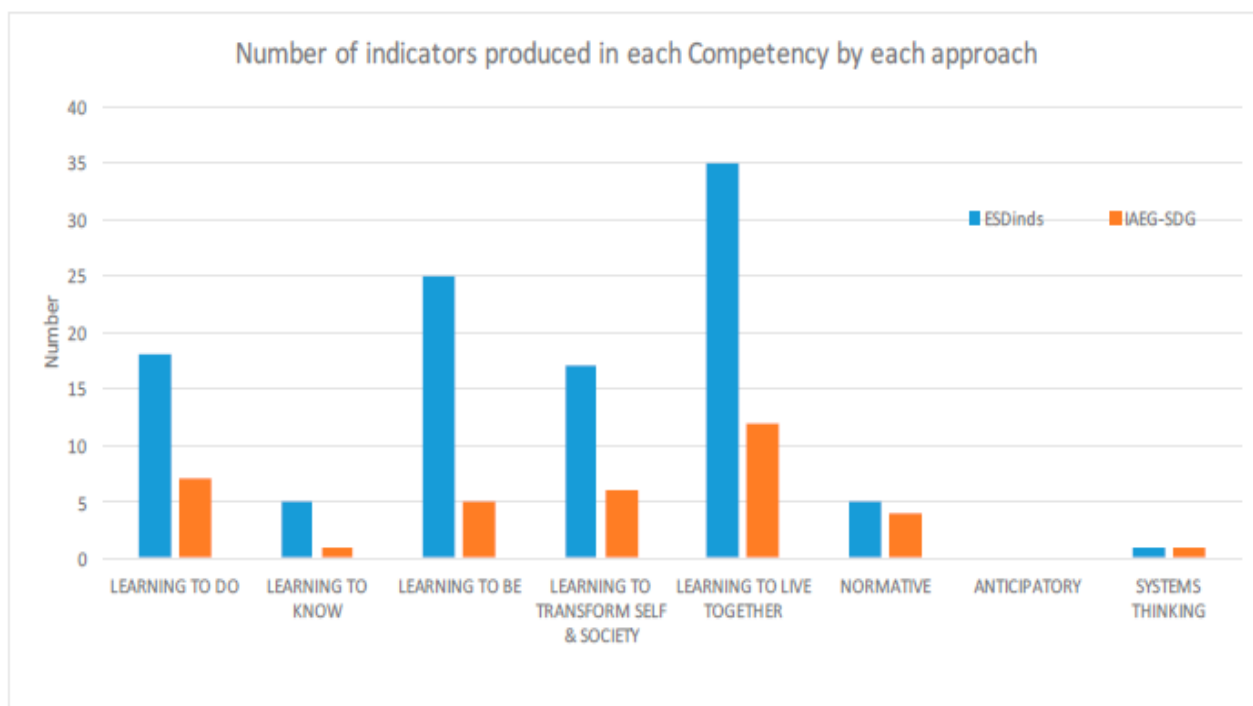
4.4.2. Inside and Outside the (Traditional) 'Box'

The questions posed by the UN's SDG consultation and the ESDinds approach covered similar domains, but the UN consultation started with one proposed indicator and asked respondents to suggest alternatives—with the implication that any operationalising indicators returned would be appropriate to that given target, and that the responses would be in the context of traditional indicator types. The ESDinds project, despite having the same aim of generating indicators, began with the question of what practitioners find `meaningful, worthwhile and valuable' in relation to their ESD work. This was usually done with a focus on successful projects that they had completed in the past, and the values underlying their success. Asking specifically for new candidate indicators may predispose respondents to thinking about indicators and data collection methods that are already familiar to them; but asking specifically about what is meaningful, worthwhile and valuable to people may help them to think outside the traditional indicator 'box'. In this latter case, they may find it easier to identify less tangible dimensions of ESD, which usually fall outside the realm of evaluation and assessment (see Chapter 3). The approach brought about useful results which could help to achieve the shift, noted by McCool and Stankey (2004), from 'what can be measured, using currently available methods' to 'what should be measured'. Figure 4-2 shows that ESDinds returned many indicators in Learning to Do, Learning to Be and Learning to Transform, which are not traditionally well covered by indicators[20].

4.4.3. Tacit Knowledge and the Values-Action Gap

The ESDinds indicators proved to cover many of the less tangible dimensions of ESD. One useful way of thinking about this result is that values elicitation, especially when based around reflection on collective projects that have been successfully completed, may provide access to tacit knowledge that has not previously been expressed in words. Tacit knowledge accumulates through shared contexts of practical experience (Polanyi, 2009; Sanders & Stappers, 2012) but may never be formally articulated unless—as in the ESDinds project and

---

[20] The original paper also included a phrase cloud, created by a co-author, which provided a visual representation of the nature of the indicators created by the two methods - using larger font sizes to represent competencies with more indicators, and smaller font sizes to represent those with fewer.

more recent adaptations of its methodology—it is actively elicited. The accumulation of tacit knowledge provides one explanation for a *value-discourse gap*—a situation in which people are already enacting a particular value in their day-to-day actions, but do not bring it into conversation (see Chapter 5 below). A 'value-discourse' gap is the converse of the previously recognised 'value-action gap' (Blake, 1999), also known as an 'environmental values/behaviour gap' (Kennedy, Beckley, McFarlane, & Nadeau, 2009) or 'attitude-action gap' (Kollmuss & Agyeman, 2002), in which people espouse a particular value in their discourse but fail to carry it forward into action (see also Rescher, 1982; Schlater & Sontag, 1994). The ESDinds approach has shown it can produce a large number of indicators covering areas which are generally considered to be much less tangible (see Appendix C for examples), and may make accessible tacit knowledge about the types of skills that are required for sustainable development. Such knowledge might not be foregrounded in traditional indicator development consultations, because of the existence of a value-discourse gap in the respondent organisations: aspects that are important in practice may not always be articulated in words.

4.4.4. Current SDG Discourse

The concept of value-discourse gaps takes on a greater significance if we regard the IAEG-SDG public consultation as a snapshot of contemporary global political discourse around the SDGs, and the Target 4.7 responses as an indication of how ESD is being discussed within the 'SDGs community'. Remembering that the IAEG-SDG responses included 22 with relevance to Sub-Target 1 (Knowledge), it is noteworthy that only one was found to be relevant to the broader 'ways of knowing'- such as critical thinking, or applying learning to life—described in the Skills, Learning to Know competencies. It seems that even though the target is about ESD, the SDG community is still focusing on the more familiar domains of the cognitive 'learning to know', e.g., in the sense of receiving knowledge, rather than challenging received knowledges or producing new systemic knowledges. Such traditional knowledge is characterised by an outward focus and readily measurable outcomes, e.g., those tested through examinations. The more affective and inward-facing dimensions of 'learning to know'—as well as many other sustainability competencies like 'learning to live together', 'learning to be' and 'learning to transform self and society' and the 'normative' competencies

that allow people to explore questions of values and ethics, and to critique prevalent definitions of SD itself—are frequently overlooked. (I discuss this point further in Chapter 5 below; see also Clugston (2011); Hedlund-de Witt (2011); Dahl (2012). This is evident in the responses obtained through the IAEG-SDG approach, itself representing current discourse. This should not be taken to mean that such competencies are not valued, but rather that people's individual tendencies to value them do not yet appear to have been translated into targeted conversations around how to facilitate their acquisition through education, and thus certainly not how to measure the success or failure of such efforts.

The ESDinds approach was designed to address this issue, and the indicators produced suggest its success. A case in point is L2B_SEL, "Develop one's personality, self-identity, self-knowledge and self-fulfilment", which had seven indicators proposed from ESDinds in comparison to the two from the IAEG-SDG approach. Thus, values-based approaches to indicator development may both support and promote an expansion of the global discourse on ESD to include these 'inner', 'heart-centred', or even 'spiritually-oriented' aspects that have eluded traditional approaches to indicator development. Such an expansion has been previously advocated in principle, although not demonstrated in terms of indicator development, by authors in both education (Sterling, 2001, 2004; A. E. J. Wals & Jickling, 2002) and sustainable design (Fuad-Luke, 2007; Walker, 2006, 2007; Wood, 2007).

The absence of indicators relating to 'learning to know' and systems thinking in the IAEG-SDG public consultation is a matter of concern for two reasons: first, because the consultation can be seen as a reflection or indication of the current state of discourse, and second, because of its implications for the future. In terms of the state of discourse, it infers that meta-learning—i.e., having an awareness of the phenomenon of learning in general and of one's own learning habits in particular, and taking control of one's learning (J. B. Biggs, 1985; Meyer & Shanahan, 2004)—appears not to be a salient issue for the respondents. In a similar way, it suggests that systems thinking may not yet have been embedded in the way in which these respondents approach complex sustainability problems. In terms of implications for the future, the likelihood is that if an issue is not covered by the consultation responses, it is unlikely to be represented in the final indicator. Bearing in mind the old adage that "what gets measured gets done" (Henshaw, 2006), this would constitute a missed opportunity to embed a more radical shift in modes of education—from a transmissive to a transformative paradigm (Jickling & Wals, 2008; A. E. J. Wals & Jickling, 2002)—and to push schools, colleges

and universities to produce graduates who are not only critical and creative 'systems thinkers', but also effective lifelong learners. This, in turn, has serious implications for the future of sustainable development, in that knowing about SD in a theoretical sense is very different from having the skills and motivation to take action, which in turn is not the same thing as actually *taking* action.

4.4.5. The Need for Localisable Indicators in ESD

Whichever indicator(s) the IAEG-SDG might ultimately select to operationalise Target 4.7, there will be a subsequent question of data collection that needs to be considered carefully. At one extreme, data for most indicators could be collected through a national-level survey, or even a simple 'tick-box' exercise within the respective government departments. This could easily become the default option, especially where resources are limited. At the other extreme, local institutions (such as teachers' colleges, universities and schools) could take the indicator as a starting point for a variety of initiatives designed to report on the achievement of the SDG target within their day-to-day practice. These could include, for example, creating new assessment criteria and rubrics for assessing learning outcomes at the individual or group level [29]; auditing existing curricula, internal policies, organisational structures and codes of conduct in relation to the indicator; and/or evaluating the extent to which these internal policies are implemented, or guidelines adhered to, in practice.

The authors would argue that while it is important for the final indicator to be standardised to the point of comparability across institutions and countries, it must also allow for localisation in its application if it is to be perceived as relevant (a condition for engagement) and produce locally valid measures. In addition, local institutions should be strongly encouraged to develop their own initiatives to set up localised operationalisations, especially if the final indicator does not fully operationalise the skills dimension—as seems likely, on the basis of the responses. Without targeted action, there is a high risk that global conversations about ESD will end up focusing primarily on the transmissive model of education—focusing on acquiring particular kinds of knowledge [48]—and that the opportunity to embed transformative, competency-based approaches could be missed.

It may be helpful, in light of this discussion, to revisit Figure 4-1—the conceptual framework created in this paper. We propose this conceptual framework as a tool in its own

right for use by local institutions (e.g., universities, schools, and civil society organisations) in their efforts to develop locally-valid ESD indicators, and apply them to the evaluation of organisational structures, policies, curricula, teaching practices and learning outcomes. The authors are currently developing new work in this area.

4.4.6. Salience and Emerging Breadth of Concepts

The existence of diverse candidate indicators or proto-indicators for a particular category, within a given context, may suggest that the category has been widely thought about and discussed within that group of people. In the ESDinds case, for example, we have noted that the category "L2D_ACT: Be an actor as well as a thinker" was used to code 17 different proto-indicators, in comparison to only three in the SDGs set. It is unsurprising but encouraging that this category is highly salient to the respondents, who are practitioners promoting ESD within the non-formal education sector—specifically in civil society organisations—within project contexts.

Other possible explanations for multiple indicators in a given category are that the category itself is difficult to differentiate from other categories within the framework; that it is too broad to be useful; or that the indicator development process is biased in favour of that category. In the case of L2D_ACT, for example, there is a significant overlap with two other categories whose wording inherently implies a shift from the realm of thought to the realm of action, namely "L2L_PAR: Participate and co-operate with others in increasingly pluralistic and multi-cultural societies" and "L2T_FOO: Promote behaviours and practices that minimise our ecological impact on the world around us". However, even among those uniquely coded as L2D_ACT within ESDinds, there are several different manifestations. These include, for example, behaving in a way that is consistent with one's words; actively helping the organisation to meet its goals; investing one's own time and resources in activities that benefit the environment or society; and being motivated and productive in one's work. This may suggest that there is a need for further conceptualisation of what is meant by "being an actor as well as a thinker", which can most usefully be done at the local level. Such an observation reflects the point that the ESDinds items are proto-indicators: they serve as triggers or prompts for localised processes of reflection, conceptualisation and

operationalisation, leading ultimately to indicators that are both salient and measurable at the local level.

The category with the highest overall number of indicators within the IAEG-SDG consultation dataset was "L2L_DIV: Tolerate, respect, welcome, embrace, and even celebrate difference and diversity in people". In this case, the multiplicity of indicators is attributable largely to the wording of Target 4.7 itself, which explicitly references this competency: "Ensure that all learners acquire the knowledge and skills needed to promote SD, including, among others . . . appreciation of cultural diversity". Thus, any indicators which replicated the wording of the target automatically scored a point for L2L_DIV. However, the same category was also associated with multiple indicators for ESDinds, where a number of different practical manifestations of welcoming, embracing and celebrating diversity were identified—e.g., accommodating different learning styles, giving people flexibility to do things within agreed boundaries, or providing different groups with equal access to information and decision-making. This highlights the possible need for further conceptualisation of L2L_DIV with reference to the literature on, and practice of, equality and diversity in both education and employment settings. As before, this may be most appropriately done at a local level.

4.4.7. Whose Targets? Whose Indicators?

Regardless of whether explicitly values-based approaches are used, our findings also open up a broader debate about indicator development in relation to the SDGs. Revisiting the United Nations call to "measure what we treasure" (United Nations, 2015a, p. 10) in the light of this study, it becomes evident that what is not discussed is the meaning of 'we' in this statement. It appears to imply that a single set of values are being used as the benchmark for determining what is 'treasured', thereby setting a normative standard for indicator design—which seems to have gone unchallenged in this instance, but has previously been the subject of critique (Lele & Norgaard, 1996; McCool & Stankey, 2004). A more generous interpretation might suggest multiple definitions of 'we', and hence a nuanced and multi-subjective / intersubjective understanding of values and indicators—echoing the position of Ostrom (2010) on the need for polycentric approaches, as well as Saith's (2006) critique of the Millennium Development Goals (MDGs). However, neither the MDGs Report nor any subsequent document relating to measurement and assessment for the SDGs incorporates

an explicit call for polycentric, decentralised approaches to SDG indicator development. Rather, as for the MDGs, the official process has been oriented towards identifying a single global indicator (or, if absolutely necessary, two or three indicators) for each of the 169 SDG targets.

Bell and Morse (2011) highlight the 'tyranny of methodology' which is inherent in any attempt to reduce a multiplicity of stakeholder perspectives to a very limited set of global indicators. Whatever is ultimately produced by the IAEG-SDG as the indicator (or set of indicators) for Target 4.7 at a global level, we speculate that it is unlikely to satisfy any of its observers. Furthermore, we have illustrated in this paper that it is impossible for a single indicator or even a handful of indicators to provide valid measures of progress towards such a broad, multi-faceted, and intermeshed target, in the sense of covering the whole of the content domain (Onwuegbuzie & Johnson, 2006; Onwuegbuzie et al., 2007).

4.4.8. Democratisation and Pluralism in Indicators

In highlighting the challenges of operationalising complex concepts such as 'education for sustainable development', we neither question the significance of the IAEG-SDG indicators and indicator design processes, nor the importance of global and national monitoring in a more general sense. Our intent is, rather, to suggest the relevance of supporting democratisation of the indicator development process. This would mean acknowledgement that designing indicators of progress towards complex goals like the SDGs is best served by complementary processes involving both expert groups and international institutions and by local values and priorities at a variety of different levels.

The inductive development of indicators from what groups identify as immediately valuable, worthwhile and meaningful in their diverse contexts would have important implications for engagement in and the democratisation and diversification of global sustainability and development agendas as a whole. The inclusion of such indicator development processes in international conversations could draw increasing attention to diverse ways of seeing and understanding the world—leading, potentially, to the wider recognition of plural 'dynamic sustainabilities' (c.f. Leach, Scoones, & Stirling, 2010)in place of the IAEG's single, static conception of sustainable development, implicit in the use of the definite article

to refer to 'the' knowledge and 'the' skills required to promote SD. This is already an acknowledged issue for Indigenous communities, in the light of their long history of disenfranchisement and ongoing collective actions for international recognition. Indeed, the importance of creating SD indicators grounded in Indigenous values and world-views was formally noted in a 2006 technical report of the Permanent Forum on Indigenous Issues (UN-PFII, 2006). However, a trend towards democratisation and diversification could also open up the possibility of creating a place at the SDG table for other sustainability discourses that do not sit comfortably with the implicitly neoliberal concept of sustainable development (Robinson, 2004). These include, for example, the interface of the arts and sustainability (Kagan, 2011); work on 'design for sustainability' positioned in opposition to 'design for sustainable development', which demands a fundamental shift in values rather than a mere 'technical fix' of contemporary industrial societies (Walker, 2006, 2007; Walker & Giard, 2013); Stephen Sterling's positioning of 'sustainable education' in opposition to 'ESD' (Sterling, 2001), grounded in similar concepts and arguments to those of 'design for sustainability'; and the ecological philosophy of David Abram (1996), which posits that members of the 'more-than-human' community of life should be treated as perceiving subjects rather than as objects for human consumption.

Rather than advocating for the immediate dissemination and use of the specific proto-indicators created through the ESDinds project, our concluding action is to highlight the pragmatic usefulness of inductive values-based approaches in operationalising ESD targets at the project level, and—in light of this—to call for increased and wider engagement in national, regional and local initiatives generally, in order to complement the official IAEG-SDG process. In view of recognised difficulties in defining terms such as 'global citizenship' in a global way (and assuming that the same competencies will be equally important everywhere), we propose that the key point is perhaps the process of creating values-based indicators, as part of a global trend towards the democratisation of the SDGs.

4.5. Conclusions and Recommendations

A conceptual framework was constructed for SDG Target 4.7 and used to analyse the usefulness and character of indicators produced from a values-based approach compared to the UN approach. The findings suggest that the current SDG community, and within it those

focusing on ESD, still conceive of or model knowledge in a narrow, traditional way relating to content—to the extent of not being able to suggest relevant indicators for different ways of knowing, such as critical thinking or 'learning to learn'. Similarly, indicators for many of the less tangible areas of ESD which are not already easily covered by traditional indicator approaches were not produced by the ESD community involved in the UN's IAEG-SDG process. Although this is likely to be a reflection of current discourse in that community, it is also likely to be a reflection of the IAEG's consultation process, which appears to have difficulty generating non-traditional indicators. However, the consultation process did produce candidate indicators which partially covered most of Target 4.7.

The ESDinds approach produced complementary and occasionally better coverage of Target 4.7. It was specifically designed to produce indicators covering the less tangible areas of ESD, and the work presented here provides evidence that it did so, as well as opening up a discourse which included finely-developed concepts in several of the less tangible areas. Although the underlying premise in the ESDinds approach is to drive the process by considering what is locally valuable, worthwhile and meaningful, it is not clear to what extent its other aspects—e.g., its intersubjective and slightly disruptive elements, its purposeful contextualisation at group level, and its consideration of values in a holistic manner—contribute to those results. However, it seems likely that all three of these elements do contribute, because the resulting indicators produced included several representing group values in action, aspects not usually articulated, and aspects often considered intangible—which were not produced in the IAEG-SDG process.

Although it may initially seem unlikely that the ESDinds approach could contribute in practical ways to an indicator development program for complex targets, such as that of the IAEG-SDG, there are in fact some ways in which it could. Firstly, if the ESDinds-type approach were run by a small number of groups prior to wider consultation, it could produce candidate indicators (or even Targets) of those rare categories, which could then be used as prompts—both for the expert groups and wider public. Secondly, the IAEG-SDG process might be enhanced if contributors were advised to develop their candidate indicators through 'disruptive' or at least intersubjective, consultations at group level, rather than relying on single representatives of organisations. Lastly, it appears that even the ESD community which is cognizant of less traditional output types such as competencies (compared to narrow knowledge) did not propose related indicators, suggesting a hegemony of sorts which hinders

untraditional measures or evidence. To get around this, the IAEG-SDG could encourage novel or non-standard ideas like those reflecting 'heart' or 'culture', or preferably produce examples of them for consultation. It certainly would need to try to include them when deciding on Targets and indicators, as they may be rare but they are valid—as made evident in our conceptual framework of Target 4.7. Finally, the ESDinds type of approach of developing indicators from the shared values of a group, but articulating them in individualised action-related and context-related examples, may serve to relieve the tension between different indicator types. The ESDinds indicators are slightly generalised versions of specific examples that are important at the group level, and may be transferable to other groups, who can then localise them in their own context. The recommendation from this for future indicator development is to build in a 'fuzzy framework' of indicators, which allow localisation. For example, the indicator "People are not afraid to make mistakes" can be scored using evidence deemed locally to be relevant. Furthermore, whatever final indicator is constructed by the IAEG-SDG for Target 4.7, the authors would suggest that asking groups around the world to immediately consult and develop their own local interpretations and scoring systems as this would not only lead to more valid and useful indicators, but would also build awareness and capacity in ESD that produced a wealth of ideas to share with less creative groups. This could become an important part of the process of increasing ESD itself.

Lastly, the work presented here makes it clear that much more research is needed to explore the ESDinds process to better understand how it produces measures of that long-sought 'intangible': sustainability. Such work is likely to reveal new knowledge not only for sustainability, but also for other linked disciplines.

Supplementary Materials: The following are available online at www.mdpi.com/2071-1050/8/9/861/s1, Table S1: Full coding of responses to IAEG-SDG public consultation on indicators for Target 4.7 of the Sustainable Development Goals; Table S2: Full coding of ESDinds/WeValue indicators.

Appendix to Chapter 4: Creating the Conceptual Framework for Target 4.7

This section provides further detail on the creation of the conceptual framework utilised in this article, to supplement the information provided in the main text.

Top Level: Identifying Sub-Targets

By analogy with the example provided by Hak et al. (2016), the first stage of framework development is the breakdown of the target into sub-targets. As noted in the main text, this cannot be achieved simply by splitting it directly into two blocks (*all learners acquire the knowledge needed to promote sustainable development*, and *all learners acquire the skills needed to promote sustainable development*), as the wording of the target poses some inherent conceptual challenges. These are as follows:

a)   The use of the term 'all learners' is problematic in operational terms because, in conjunction with Goal 4's call for "lifelong learning opportunities for all", it effectively implies assessing the entire population of the planet. To address this, we focus on individuals who are engaged in some form of organised learning activity, whether formal or non-formal. We acknowledge that in an ESD context the domain of 'non-formal education' may include activities in which the learning element is implicit, such as reforestation projects and wildlife surveys, as well as more structured learning activities such as 'Forest Schools' and 'Wildlife Clubs' (Harder, Velasco, et al., 2014).

b)   By focusing on learning outcomes, we are implying that educators must themselves possess relevant knowledge and skills and be encouraged to put them to appropriate use, in order to facilitate learners to acquire them. This also applies, to some extent, to individuals who manage educational institutions and develop curricula.

c)   The wording of the target is ambiguous in relation to the long list of content areas—human rights, gender equality, etc. It is apparent, however, that these are not intended to be ends in themselves, but means through which knowledge and skills might be acquired. The various content areas are suggested to be necessary but not sufficient for the acquisition of the

knowledge and skills required to promote SD, as conveyed by the phrase "among others". It may be the case that there are other necessary content areas, or that the overall picture is a holistic one, in which the 'overlaps' between the various content areas—which may ultimately be undefinable—constitute the space in which the most transformational learning occurs. There may also be ways of acquiring the relevant knowledge and skills *without* passing through any of the eight specific content areas.

d)   The use of the definite article (*the* knowledge and *the* skills needed to promote sustainable development) implies the existence of a discrete, bounded body of knowledge and a clearly definable set of skills, in turn relating to a single unambiguous concept of sustainable development. As this position contradicts any reasonable representation of the state of agreement on these points, we would dispute the inclusion of 'the', and have omitted it from our proposed sub-targets. It has been widely acknowledged, for example, that the definition of sustainable development is a 'wicked' problem, characterised by "a lack of clarity, uncertainty, ambiguity, high risk, and limited understanding" (Pryshlakivsky & Searcy, 2012, p. 100).

Taking all of these points into consideration, we proposed two sub-targets for operationalisation as outcome indicators for Target 4.7:

Sub-Target 1: All learners and educators involved in organised learning activities, both formal and non-formal, acquire knowledge needed to promote sustainable development

Sub-Target 2: All learners and educators involved in organised learning activities, both formal and non-formal, acquire skills needed to promote sustainable development

Second Level (Skills Dimension Only): Broad Competency Domains

As discussed in Section 4.2 above, we have chosen to develop the second level of the conceptual framework with reference to literature on 'sustainability competencies', also referred to as 'ESD competencies'. The term 'competency' has been defined by Rychen and

Salganik (2001, p. 51) as "the ability to successfully meet complex demands in a particular context through mobilisation of psychosocial prerequisites (including both cognitive and non-cognitive aspects)", and as "a complex action system encompassing cognitive skills, attitudes, and other non-cognitive components".

We have selected two systematic models, namely the UNESCO/UNECE model (derived from the four 'pillars of learning' of the Delors Report (1996) with the subsequent addition of 'learning to transform oneself and society' (Combes, 2005), as detailed in Section 2.1, and the model developed by Wiek et al. (2011) on the basis of a systematic review of academic literature. As discussed above, these models were chosen on the grounds that they collectively encompass most of the other literature on ESD competencies, including more recent work. In combining them, we note that the Wiek category of 'interpersonal competencies' can be seen as largely equivalent to the UNESCO/UNECE 'learning to live together', and 'strategic competencies' as equivalent to 'learning to do'. Thus, we have condensed these into a single category in each case, creating an eight-category rather than a ten-category conceptual framework.

Third Level: Specific Clusters of Competencies

To create the third level of the framework, we utilised a UNESCO infographic (Oladottir, 2014) in which the five broad competency domains are broken down into specific clusters of competencies. Thus, for example, the domain of 'learning to live together' incorporates competencies in coping with situations of tension, violence and conflict; accepting and celebrating diversity and difference; responding constructively to cultural diversity and economic disparity around the world; participating and co-operating with others in a multi-cultural society; and developing an understanding of other people and their histories, cultures, etc. These, which we refer to as 'sub-domains', are shown in full in Figure 4-1. For the purpose of our framework, we assigned each of them a unique code, consisting of a designator for the domain as a whole (e.g., 'L2L' for 'learning to live together') followed by a three-letter code for the sub-domain itself.

It is important to note that there is considerable overlap between the sub-domains in the UNESCO/UNECE framework, and it is not always meaningful to attempt to differentiate them. This lack of specificity is relevant to our analysis in that an indicator may be found relevant to more than one cluster, not because the indicator itself lacks precision, but because

the clusters to which it is relevant are poorly mutually distinguished. Notwithstanding these concerns, we regard the model as a useful initial approximation of a detailed conceptual framework for ESD competencies, which can be further enhanced and clarified by future work.

Another key issue is that some of the items in the framework may be contested or controversial, such as the sub-domain 'Acquire universally shared values' under the pillar/theme 'Learning to Be', which evokes challenging questions. Which values are deemed to be universally shared? Who is (or should be) responsible for making such decisions? What happens (or should happen) when the values of a specific individual or group come into conflict with the 'universal' set? In this paper, we have opted to evade this controversy by replacing the problematic 'shared universal values' with the less sensitive 'common global values'. This latter term relates to the analysis by Burford et al. (Chapter 2), which illustrates references to global values in different United Nations documents.

Chapter 5:

Promoting sustainability skills in schools: designing tools to bridge the gaps between values, discourse and action

Abstract

It has often been observed that even when people publicly espouse certain values, they do not inevitably perform the actions or behaviours that one would expect to be associated with these values. This has been termed a 'value-action gap'. Academic research on the barriers to pro-environmental behaviour has served primarily to highlight the complexity of this area; but a problem-centred approach to learning, led by civil society organizations, has been shown to generate effective solutions. One example is the design and use of values-based indicators – statements that link generic or specific `values vocabulary' to specific real-world referents such as behaviours or perceptions.

In this chapter, we discuss the application of values-based indicators for the twofold purpose of reflection (inspiring teaching and learn-ing) and evaluation (guiding organizational development). We first describe the EU-funded project within which values-based indica-tors were initially developed, and provide an overview of the pro-cesses leading to the initial design of a project evaluation toolkit ('WeValue') and the evidence of its usefulness for identifying and bridging value-action gaps in civil society organizations providing education for sustainability. The central section of this paper reports on a co-design project to develop a toolkit of values-based indicators for secondary schools, primarily for teaching and learning purposes. Finally, in the discussion section, we suggest a theoretical grounding for the use of values-based indicators to close value-action gaps; identify a new kind of gap that has not previously been described in the literature; and reflect on some of the wider implications of our work.

Chapter 5. Promoting sustainability skills in schools: designing tools to bridge the gaps between values, discourse and action

5.1.   Introduction: Learning to 'walk our talk'

'Values' have been adopted by innumerable civil society organizations (CSOs), governments and businesses worldwide as a way of articulating their goals for ethical and sustainable practices. Although there is no universally accepted theoretical definition of values, some of the most influential definitions are those proposed by Kluckhohn, Rokeach and Williams, which describe them respectively as conceptions of 'the desirable' (Kluckhohn, 1951); enduring beliefs that a certain behaviour or condition is preferable to its opposite (Rokeach, 1973); and "criteria or standards of preference" with cognitive, affective and directional aspects (Williams, 1979, p. 16).

In accepting these definitions of values as beliefs about what is desirable or preferable, one might anticipate that an individual's values would invariably be 'enacted', i.e. manifested in their actions on a day-to-day basis (c.f. Meglino & Ravlin, 1998). Paradoxically, however, it has often been observed that even when people publicly espouse certain values, they do not inevitably perform the actions or behaviours that one would expect to be associated with these values. This has been termed a 'value-action gap' (Blake, 1999), or, in lay terms, a failure to "walk the talk" (Kennedy et al., 2009). The terms 'attitude-action gap' (Kollmuss & Agyeman, 2002) and 'environmental values/behaviour gap' (Kennedy et al., 2009) have also been used in academic literature. In this chapter, we use the broader term 'value-action gap' to reflect widespread usage, and in accordance with the observation that the gaps themselves may occur more at the level of specific actions than long-term behavioural trends (Kollmuss & Agyeman, 2002).

There is a complex and extensive literature on the factors (other than values) that influence behaviour, and the various barriers that may prevent people from undertaking specific pro-environmental actions even when these are congruent with their values (see, for esample, B. Brown, Buchanan, DiSalvo, & Margolin, 2014; Fontaine, Poortinga, Delbeke, & Schwartz, 2008; Jackson, 2005; Kennedy et al., 2009; Kollmuss & Agyeman, 2002; Patten, 2013). Perhaps because of this complexity, which may be virtually impossible to condense

168

into a single framework or model (c.f. Kollmuss & Agyeman, 2002), there is a notable absence of empirical studies that present workable solutions to the problem of value-action gaps – especially in a formal education context. In this chapter, we do not attempt the challenge of identifying, and proposing strategies for removing, each separate barrier to pro-environmental behaviour (or, more broadly, to actions that foster sustainable and responsible living). Rather, we present a holistic solution that has already proven helpful for identifying and bridging value-action gaps within the context of civil society organizations (CSOs) – namely, the use of peer-elicited values-based indicators to stimulate collective reflection (Burford, Hoover, et al., 2013; Burford, Velasco, et al., 2013; Harder, Velasco, et al., 2014; Podger et al., 2013)– and demonstrate that this solution can be adapted for use in schools.

Although a systematic review of barriers to pro-environmental behaviour is beyond the scope of this chapter, we outline some important findings from this literature to set the scene. We then describe the processes leading to the creation of a toolkit of values-based indicators ('WeValue') through an EU-funded project, and illustrate its usefulness to CSOs providing education for sustainable development (ESD) or education for sustainable and responsible living (EfSRL). The central section of this paper reports on the application of the 'values-based indicators' approach to EfSRL in secondary schools, through the design of modified toolkits for teachers, students and school administrators. Finally, we relate this new work back to the literature on value-action gaps and identify a new type of 'gap' that has not previously been discussed, as well as reflecting on the implications for design literature.

5.2.   Value-action gaps: brief overview of relevant literature

In the early days of environmental education, it was assumed that an educator's sole task was to instil knowledge of specific environmental problems, and that desirable pro-environmental behaviours would follow automatically. This assumption has since been demonstrated to be fallacious (B. Brown, Buchanan, DiSalvo, & Margolin, 2013; Heimlich, 2010; Kennedy et al., 2009). Furthermore, rational-choice theories of decision-making such as the Theory of Planned Behaviour (Ajzen, 1991), which suggested that individual actions are the result of conscious cognitive deliberation of values, attitudes and social norms, have been largely discredited.

Criticising the Theory of Planned Behaviour, Jackson (2005) particularly highlights the importance of entrenched habits, routine, instinct and emotion in influencing human behaviour, and notes that situational factors may make specific value orientations more salient than others at certain times (see also Peng et al., 1997). The latter point is significant because personal values, when held at a subconscious level, can be mutually contradictory: while acting in accordance with some of their values, individuals may be forced to violate other values (Kennedy et al., 2009; Redclift & Benton, 1994). Thus, what appears as a value-action gap could, instead, be attributable to what might be termed 'over-ruling' of one value by another: for example, a person who holds strong pro-environmental values, but also values frugality, might ultimately refuse to purchase expensive organic food (Kennedy et al. 2009). Kollmuss & Agyeman (2002, p. 250) suggest that even when a person's broad lifestyle choices are based on altruistic and social values, their motives for specific actions are often more selective and revolve around immediate needs: comfort, money and time.

Diverse studies reviewed by Jackson (2005) have demonstrated that 'over-ruling' can be manipulated, e.g. by framing situations in a particular way or priming certain values through the use of appropriate images. This is possible because, under normal circumstances, values are "less than totally conscious, somewhat below an individual's level of complete awareness" (Meglino and Ravlin 1998, p. 360; see also Rokeach, 1985; Kopelman, 2003). They have been described by Goleman (1998) as "intimate credos that we may never quite articulate in words so much as *feel*". Thus it is possible to strengthen particular values precisely by attempting to articulate them in words, e.g. by reflecting on one's reasons for espousing them, thereby drawing them out from the affective realm into the cognitive realm. As Maio et al. (2001, p. 14) explain: "We believe that…generating reasons for a value provides concrete examples of why behaving consistently with the value is sensible and justified. Thus, when situational forces work against provalue [sic] behaviour, people become able to retrieve concrete information in addition to their vague feelings about the value."

These findings are consistent with observational research conducted in a real-world educational setting more than four decades ago. Dixon (1978) observed that providing 'values clarification' exercises to children, which effectively sensitised them to the values that they already held (c.f. Raths, Harmin, & Simon, 1978) could reduce confusion and apathy and increase desirable classroom behaviour (Schlater & Sontag, 1994).

The phenomenon of value-action gaps is not limited to the individual level, but also has important implications for organisational learning and behaviour. It is often informative to take a group (i.e. an organization or project), rather than the individual, as the level of analysis (Agle and Caldwell 1999; Meglino and Ravlin 1998). Bansal (2018)) has illustrated, for example, that environmentally responsible action is more likely to be taken when it is consistent with both individual concerns and organizational values. Conversely, research into corporate social responsibility and sustainability policy adoption shows that many organizations face discrepancies between formal commitments and actual policy implementation (Wilber, 2004). Accountability for adherence to espoused values, through the adoption of measurable indicators, may provide a means for overcoming this disconnect (Gruys et al. 2008). In the next section, we describe a novel approach to values clarification through reflection on 'indicators' that can help EfSRL-promoting schools to create shared understanding around the enactment of their espoused values.

5.3. The ESDinds Project and WeValue evaluation toolkit

The ESDinds project, funded from 2009-2011 by the European Commission's Seventh Framework Program, brought together CSOs and academic researchers from five countries to collaboratively develop useful indicators and assessment tools to evaluate the 'presence' and enactment of core values. The project aimed to establish values-focused evaluation and reflection within a diverse range of businesses and civil society organizations (CSOs), especially those promoting EfSRL (Harder, Velasco, et al., 2014; Podger et al., 2010). The research design for this process is outlined in Figure 5-1.

Figure 5-1. Process of developing values-based indicators and assessment tools within the ESDinds project. Adapted from ESDinds (2011).
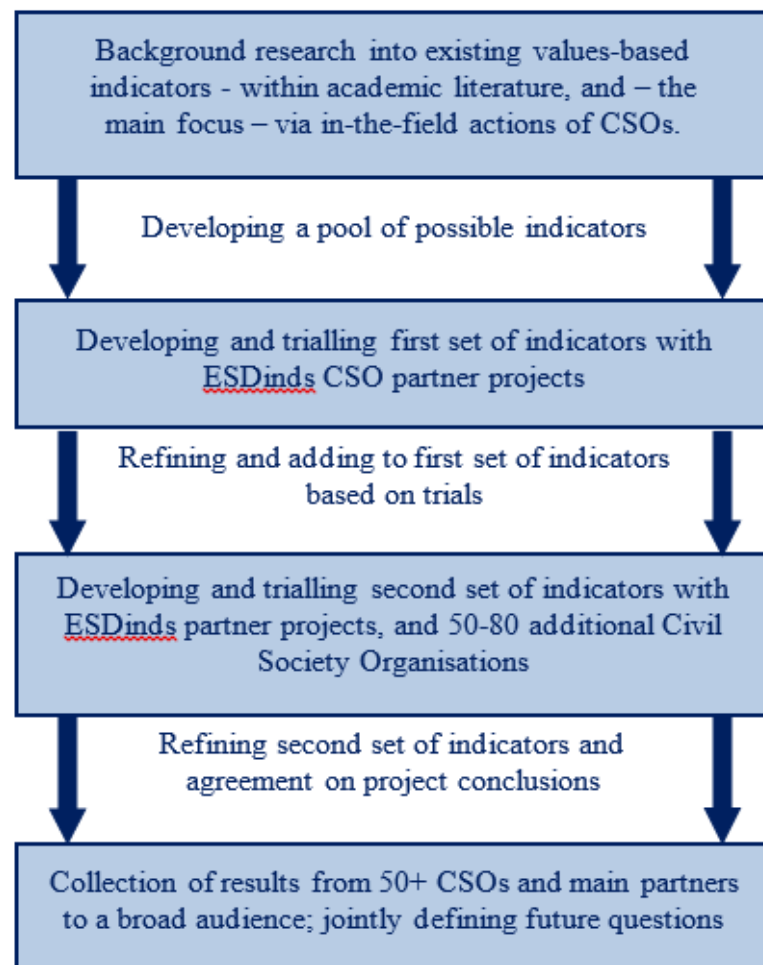
This iterative and grounded approach to indicator development led initially to the creation of a set of peer-elicited `indicators' for six specific, named values that were important to the CSOs – i.e. trust, integrity, justice, empowerment, unity in diversity, and care and respect for the community of life (Podger et al., 2010; ESDinds, 2011). However, the field testing revealed that the value-indicator relationship was not straightforward, and what had initially been seen as indicators of values were perhaps more accurately described as 'proto-indicators' – statements of ideal realities that may not in themselves be measurable, but can provide starting points for the development of measurable indicators within defined local contexts (Harder, Velasco, et al., 2014). The final phase of field testing led to a more flexible set of values-based proto-indicators, presented as a single list without specific value headings (Burford, Velasco, et al., 2013).

The ESDinds Project has also led to the co-design and testing of WeValue, a toolkit aimed at helping organizations (especially those providing or promoting EfSRL) to clarify shared values and develop context-specific evaluation strategies (see Chapter 3). Recent unpublished research has distinguished five interlinked and iterative steps in the process:

(i)   values elicitation (individual and/or collective reflection from scratch on what participants find meaningful, important and worthwhile within their project or organisation), often with a user-selected or co-created image as the starting point;

(ii)   individual and/or collective reflection on a reference list or 'menu' of values-based statements / `proto-indicators';

(iii)   using the results of steps (i) and (ii) to create specific, measurable indicators for the enactment of core values within the local context;

(iv)   identifying appropriate assessment methods that can be used to evaluate the project in relation to each of the chosen indicators;

(v)   collecting and analysing relevant data.

This approach encourages a localized 'dialogue of values' (Maturana & Varela, 1991) around sustainability-related goals and actions, addressing values discourse as well as associated behaviours. As illustrated in the examples below, the exact nature, sequence and relative importance of steps in this process will differ from one organization to another. The reflective steps (i) and (ii) may take precedence over the evaluative steps (iii) to (v), especially where the primary purpose is teaching and learning, rather than evaluation or assessment for its own sake. We have recently described, for example, a study in which values-based indicators were used as a tool for assisting undergraduate students to improve their group work skills (Burford et al., 2014). In this study, although student groups were asked to choose indicators for reflecting on their performance in different tasks, there was no formal grading: the key outcome was meta-learning.

5.3.1. How might the WeValue toolkit help users to close value-action gaps?

The WeValue toolkit has been extensively tested in real evaluation contexts in diverse civil society organizations, spread over three continents (Burford, Hoover, et al., 2016; Burford, Hoover, et al., 2013; Burford, Velasco, et al., 2013; Harder, Velasco, et al., 2014; Podger et al., 2016; Podger et al., 2013).   Reports produced by researchers and CSO staff participating in these trials suggested that the WeValue toolkit might yield other benefits, beyond the successful achievement of users' self-selected evaluation goals (see Chapter 2). Some of these benefits can be framed in terms of the identification of value-action gaps, design of possible behaviour change solutions or new actions to close the gaps, and/or implementation of those changes:

Example 1: Post-conflict youth project in Sierra Leone: "Team members organised in groups of 3–4 were asked to enact through role-play, and then to discuss, examples of discrimination and good treatment (non-discrimination) respectively in the wider communities and in their teams. They were also asked to enact ways in which the situations of discrimination might be changed, and to discuss opportunities and barriers to behaviour change." (Burford, Hoover et al., 2013, p. 7)

Example 2: Cross-faculty environmental action programme in a Mexican university: "The content of peer education workshops has moved away from a primary focus on concrete behaviours, such as recycling waste, to a holistic values-centred approach that is anticipated to generate the desired behaviours in a more deep-seated and sustained way." (ibid, p. 11)

Example 3: Small civil society organisation in Germany, using theatre-based methods to teach young people about conflict resolution: "[WeValue] brought values consciousness to the forefront of PT's activities, and strengthened its identity as a values-based organisation... Both the orientation programme for new volunteers, and

the way in which the goals of PT's work are communicated to new schools, have been restructured to centre around values." (ibid, p. 11)

Example 4: Mexican youth group promoting reforestation and arts-based activities: "The process helped them to identify values in action. Based on what values, they take what decisions? For example, one youth, 'Carlos', was a good example of 'before and after'. He is mid-way through the age range and beginning to participate a lot more. He used to be very unfocused, but after the process and specifically through the exercise, it allowed him to identify where he was. Now he participates, relates more, has more leadership." (Podger et al. 2013, p. 24)

In each case, value-action gaps were identified through individual or collective acknowledgement that a values-related 'ideal' situation (as defined by one or more proto-indicators from the ESDinds Project reference list), was not sufficiently represented within the respective organisation or project. Activities such as role-play and focus groups, designed as methods of collecting evaluative data, created safe spaces in which these gaps could be discussed openly and possible solutions explored. While in example 1 it is unclear whether this proceeded beyond the discussion and enactment stage, the other case studies provide evidence of observable organizational responses (in the form of the redesign of training activities, communication strategies and resource materials: examples 2 and 3) as well as individual responses (in the form of spontaneous behaviour change: example 4).

5.4. Closing value-action gaps in schools? Towards a usable toolkit

Taking the above insights as its starting point, a workgroup supported by the PERL project set out to design a modified version of the WeValue toolkit that might contribute to the identification and closure of value-action gaps within a secondary school context. This was based on the understanding that, as stated in the 2012 PERL work plan, "Throughout the decade, most education about sustainable lifestyles has centred around explaining the dire consequences of what has been done wrong". Such approaches are, however, often

conducive to inertia and despair: in order to stimulate active agency and achieve real change, it is often necessary to frame EfSRL in more positive and constructive terms (Harder, Burford, & Hoover, 2014). This may be achieved by co-create visions of a desirable future, to and to "examine and identify the values base from which [these visions] should spring" (PERL 2012; see also Harder, Burford & Hoover, 2014).   The design and use of a modified version of the WeValue toolkit was hypothesised to contribute towards stimulating reflection on values, and their enactment in practice, among teachers in secondary schools.

5.4.1. Epistemological and methodological approach

In working towards a values-based toolkit appropriate for secondary schools, we adopted a co-design approach, using a 'Research through Design' framing in the first phase. Research through Design can be understood as "making the right things", i.e. creating artefacts that are intended to transform the world from a current state to a preferred state (Frayling, 1993; Zimmerman et al., 2007). It can be differentiated from conventional research in both the sciences and the arts by being grounded in the specific epistemology of design described by Cross (1999), which focuses on modelling and synthesis: see Table 4.2.

Table 5-1. Epistemology of design research contrasted with those of research in the sciences and arts; adapted from Cross (1999)

	Design	**Sciences**	**Arts**
'Things to know' (fields of knowledge)	Artificial world	Natural world	Human experience
'Ways of knowing' (core values)	Imagination and practicality	Rationality and objectivity	Reflection and subjectivity
'Ways of finding out' (intellectual skills)	Modelling and synthesis	Experiment and analysis	Criticism and evaluation

In addition to being distinct from other forms of research, Research through Design is also clearly distinguishable from `normal' design practice, and from the types of research that might be conducted within the course of a commercial design activity - e.g. explorations of user experience as a precursor to the design of marketable products, often termed 'research for design' (Cross, 1995; Zimmerman et al., 2007). This distinction can be summarised in the two interrelated concepts of contribution to knowledge and contribution to society. As explained by Forlizzi (2014, p. 24): "In Research through Design, the designer seeks to understand a problematic situation in the world, and to codify that knowledge, along with a suggestion for an improved future state, in the form of a redesigned thing" (see also Frayling, 1993; Zimmerman & Forlizzi, 2008). The emphasis is therefore placed on responding to complex or 'wicked' societal problems that have no simple or clear solutions (c.f. Buchanan, 1995; Farrell & Hooker, 2013) rather than on commercial success (Zimmerman et al., 2007).

The knowledge generated by a successful Research through Design process often extends beyond the theory embodied by the artefact itself, in its particular framing of the 'preferred' versus the 'current' situation (Cross, 1999; Zimmerman and Forlizzi, 2008; Zimmerman et al., 2007) to encompass other knowledge outcomes. These may include, for example, the development of novel design processes and methods; the emergence of future research agendas, often in the form of a "nascent theory of the near future"; and the application of design to new areas (Zimmerman and Forlizzi, 2008, p. 44).

The starting point for the project described in this chapter can be summed up in the following three pairs of propositions, which collectively constituted a statement of the problematic situations prior to the start of the project and the preferred future situations (shown in Table 4.3 below).

Table 5-2: Propositions underlying this study

Problematic situation (at start)	Preferred future situation
1A: Teaching of EfSRL in schools typically focuses on examining current global problems (e.g. climate change) and their likely consequences: may contribute to apathy and despondency	1B: Teaching of EfSRL in schools focuses on developing values and skills necessary for envisioning and co-creating better futures; contributes towards a sense of power to effect change
2A: Even when students and teachers do envision `better futures', they may not recognise where their current actions and behaviours are incongruent with these futures, or take appropriate and effective steps to modify them (i.e. value-action gaps are not identified and closed)	2B: Students and teachers understand where their current actions or behaviours are incongruent with their envisioned 'better futures' (i.e. identify value-action gaps) and take appropriate and effective steps to modify them (i.e. begin to close these gaps)
3A: Although evidence suggests that a values-based indicators toolkit may be helpful in ameliorating problematic situations 1A and 2A, the available toolkit (WeValue) is not fit for purpose because it was developed with and by CSOs in a project evaluation context and its vocabulary reflects the values and priorities of CSOs, albeit with an interest in EfSRL.	3B: A values-based indicators toolkit is developed with and by teachers and students in a secondary school context, such that its vocabulary reflects values and priorities of a positive approach to EfSRL within formal education. The toolkit is effective in transforming problematic situations 1A and 1B into preferred situations 2A and 2B, respectively.

5.4.2. Developing a new 'menu' of values-based proto-indicators

In accordance with the observation that the ESDinds Project indicators reflected the values and priorities of CSOs, the purpose of the research phase was to create a new reference list of values-based proto-indicators, relevant for EfSRL teaching in secondary schools. To do this, we conducted a new analysis of several datasets that we had previously collected during our explorations of values in educational contexts:

(i)   field notes from participant observation and survey questionnaires completed by participants in an education conference in Ireland, as part of the ESDinds project;

(ii)   transcripts of semi-structured interviews with eight lecturers at the University of Brighton, including some in the School of Education;

<blockquote>

(iii) transcripts of semi-structured interviews with secondary school teachers in Tanzania.
</blockquote>

Consent had been previously been given by the participants, at the time of data collection, for the data to be used for the development of values-based indicators.

The datasets were analysed in using qualitative content analysis to identify value-labels and referents. We defined 'value-labels' as words or phrases that appeared to signify an abstract concept that was valued by the respondents, e.g. 'fun' or 'engagement'. 'Referents' were understood as direct quotes from the transcripts that referred to actions or affective states which the respondents associated, explicitly or implicitly with the enactment of these values, e.g. "see the funny or ridiculous side of the subject area", or "[students have] thought about what you've said". The value-labels were then aggregated into broader categories or themes which can be understood as clusters of values (e.g. "fun / humour / silliness" or "engagement / initiative / responsibility". The analytical process was cumulative, generating a total of 31 value clusters across the four datasets, as shown in Table 5-3.

Table 5-3: Overview of the 31 value clusters identified from the four datasets, organised in alphabetical order

THEME (VALUE CLUSTER)	UK: LECTURER INTERVIEWS	IRELAND: EDUCATION CONFERENCE FIELDNOTES	IRELAND: EDUCATION CONFERENCE SURVEYS	TANZANIA: SCHOOL TEACHER INTERVIEWS
academic excellence / examination performance			Yes (new)	Yes
challenge / risk-taking	Yes	Yes	Yes	No
community action / connection / 'real world'	Yes	Yes	Yes	Yes
compassion / caring	Yes	Yes	Yes	Yes
creativity	Yes	Yes	No	No
dialogue / collaboration	Yes	Yes	Yes	No

discipline / behaviour				Yes (new)
enabling / empowering	Yes	Yes	Yes	No
engagement / initiative / responsibility	Yes	Yes	Yes	No
extra-curricular			Yes (new)	No
financial benefits				Yes (new)
flexibility / inclusivity	Yes	Yes	Yes	Yes
fun / humour / silliness	Yes	Yes	No	No
integration / holism	Yes	Yes	Yes	No
leadership / facilitation	Yes	No	Yes	No
learning environment			Yes (new)	Yes
love / friendship / closeness				Yes (new)
'parenting' role				Yes (new)
personal goals / employment / progress				Yes (new)
peer support			Yes (new)	Yes
positivity / happiness	Yes	No	Yes	Yes
preparation / resources				Yes (new)
professional development			Yes (new)	Yes
reflection / criticality	Yes	Yes	Yes	No
relationships with parents			Yes (new)	No
respect	Yes	No	Yes	Yes
rights			Yes (new)	No
sacredness			Yes (new)	No
safety / security	Yes	Yes	Yes	Yes
self-knowledge				Yes (new)
sense of place / roots		Yes (new)	No	No
service / giving				Yes (new)
student-centredness			Yes (new)	No
transformation	Yes	No	No	No
understanding				Yes (new)

Total themes in dataset	16	13 (1 new)	22 (9 new)	20 (9 new)

Following this analysis, the referents for each theme were aggregated across the four datasets and examined as a complete set, removing duplicates and refining wording to create an initial list of proto-indicators. The criteria for defining a proto-indicator were, first, that it represents a statement of an ideal or valued reality; second, that it contains a subject, even a vague one such as 'people', and a verb; and third, that it is seen by the researcher as potentially 'measurable' or at least pointing towards something that can be evaluated, e.g. through observation, surveys, and/or qualitative methods such as interviews and focus groups. The definition of proto-indicators is a highly subjective process, but in accordance with the underlying design epistemology, we adopted a pragmatic approach to creating a usable proto-indicator 'menu' rather than attempting to represent every nuance.

The full reference list of over 300 proto-indicators was reviewed by the PERL project workgroup, consisting of seven members with a diverse range of professional roles and experiences in EfSRL. Through an iterative process of selection, clustering and discussion, it was reduced to a shortlist of 38 that were felt to be useful for evaluating schools and 15 that were felt to be helpful for supporting teaching and learning at the classroom level.

Reflection on the latter shortlist highlighted, however, that it was still not fit for purpose, in that the key proto-indicator "Students acquire values and competencies different from those of materialistic, technocratic societies" did not provide sufficient detail about what the desired competencies might actually *be*. To remedy this, the researchers identified an established ecopsychology text providing detailed information on values and skills underpinning the creation of sustainable communities, based on more than 25 years of research in diverse Indigenous societies, namely *Nature and the Human Soul* by Bill Plotkin (2009). A content analysis of selected chapters of this text was conducted to generate new indicators for review by the workgroup, and 37 of these were added to the 'teaching and learning' shortlist. Additional proto-indicators were also contributed by PERL workgroup members: some directly, and others through a written survey (modelled on the 'Educate

Together' questionnaire described above, but with a stronger EfSRL focus). After further revisions by workgroup members and three UK secondary school teachers, the final shortlists consisted of 32 proto-indicators for whole-school evaluation and 42 for supporting teaching and learning.

Following informal feedback from colleagues, teachers and young people, it became clear that while it might in principle be possible to engage teachers in reflecting on values-based indicators within their in-service training, a more immediate and appealing design prospect was a toolkit that could be used with students in the classroom. At this point, some of the indicators were reworded to make them more accessible to youth. We also realised that since the focus of the toolkit had shifted towards reflection and learning rather than formal evaluation, it would be more useful to refer to the statements as `skills for sustainable and responsible living' (SRL) than as `proto-indicators'.

Having shifted focus from teachers to students, the design of the actual activities that would constitute the toolkit itself was heavily influenced by Kim Sabo Flores's (2008) pioneering work on 'Youth Participatory Evaluation' (see also Hochachka, 2005; Seamon & Zajonc, 1998). Drawing on Vygotsky's theory that children develop and learn by "performing a head taller than they are" (Torbert, 2001), Sabo Flores highlights the importance of play and performance in youth participatory evaluation, and advocates relating to young people "as evaluators, not merely as if they were evaluators" (Sabo Flores, 2008, p. 23; see also Seamon and Zajonc, 1998). (This subtle but crucial distinction can be understood through the analogy of watching actors in a theatre 'as' their characters, rather than 'as if' they were their characters).

We modified some of the workshop activities proposed by Sabo Flores (2008) to make them suitable for values elicitation in schools:

a) "The First Thing You Think Of": asking students to write down the first thing that came into their minds when the facilitator mentioned certain words, i.e. 'participation', 'community', 'sustainability', and the name of the school itself (c.f. Sabo Flores, 2008, p. 52)

b) "The 'Yes, And...' Game": encouraging students to create a `collective story' about the type of future they would like to see for their school, in which each new participant had to acknowledge the preceding contribution by saying "Yes, and..." (c.f. Sabo Flores, 2008, p. 56).

c) 'Human survey' to assess the extent to which the students felt that key skills were already being put into practice in the school, by asking them to arrange themselves along an imaginary line across the room that represented a scale from 0% to 100% (c.f. Sabo Flores, 2008, p. 50). This has parallels with the 'spatial survey' method that we had previously tested during the ESDinds Project, in which participants were required to move into one of three different physical spaces to represent their choice from three possible answers to a question (Burford, Velasco, et al., 2013).

These new activities were included in the prototype toolkit alongside a number of established activities from the ESDinds Project, such as reflecting individually and collectively on the reference list of statements; selecting those that stand out as particularly relevant or important; grouping and prioritising the chosen statements; and reflecting on them through spatial surveys and other non-cognitive methods such as role-play.

Figure 5-2. Overview of the Research through Design process

1) Development of intial 'formal education' indicator set
- Content analysis of pre-existing datasets examining values in education contexts.
- 31 value clusters across four datasets identified, and used to identify and develop potential indicators
- OUTCOME: over 300 indicators

Expert group input

2) Expert panel review
- Expert group input to prioritise and review the indicators
- OUTCOMES: 38 indicators for evaluating schools and 15 for supporting teaching and learning at the classroom level - more specificity needed for describing values and competencies

Expert group input and informal feedback from educators and teachers

3) Further development of indicators and 'skills'
- Incorporation of data from expert group and comments from 3secondary school teachers
- OUTCOME: shortlist of 32 proto-indicators for whole-school evaluation and 42 for supporting teaching and learning.

Litereature on youth participatory evaluation and ESDinds project experience

4) Towards developing a toolkit for 'skills for sustainable and responsible living'
- Development of activities
- OUTCOME: Prototype toolkit for teaching and learning in schools (evaluation toolkit to follow)

5.4.3. Capacity building in an English secondary school

Following the 'Research through Design' phase of the project, we set out to train teachers and students in using the toolkit. We worked through the toolkit activities with nine 'student governors' (i.e. students who had been elected by their peers as members of the school council) aged between 14 and 17, in two separate sessions at an English secondary school.

The sessions with the student governors had two separate aims: testing the newly designed activities, and training these youth as peer facilitators with a view to having them subsequently facilitate activities for younger students (aged 12-13).   These exercises also highlighted some new SRL skills, which we added to the provisional 'menu' (see section 1.4.4 below).

In the first session, we began with the two values elicitation exercises described above, "The First Thing You Think Of" and "The `Yes, And…'Game". We then introduced the reference list of skills and asked the student governors to reflect on the relative importance of the skills to them, first individually (by placing green and red stickers next to those they viewed as most and least important, respectively, on their own copy of the original list) and then collectively (by using wrapped chocolates to 'vote' for their favourites on the large list, with the new skills added). The activities were well received, although the collective story was challenging for some students, and there were several awkward silences.   In discussing the skills list, it became evident that many of these are skills which students have already learned – often through arts, sports and other extra-curricular activities, as well as the core curriculum – but may not necessarily have articulated before.

The activities carried out in the second session were as follows:

a)   a silent throwing and catching activity used as an ice-breaker, to help students become more comfortable with silence and thus reduce awkwardness (also useful for general stress reduction: c.f. San Francisco United School District, 2014);

b)   grouping SRL skills that 'belonged together', using individual printed cards with one skill per card;

c)   prioritising three key skills from the collective story and/or the 'menu' list;

d)   carrying out a 'human survey' to assess the extent to which the students felt that these key skills were already being put into practice in the school, by asking them to arrange themselves along an imaginary line across the room that represented a scale from 0% to 100% (Flores, 2008, p. 50; see also Author et al., 2013a, on 'spatial surveys');

e)   enacting the chosen skills through role-play, in small groups.

The goal of testing the new activities was successfully achieved. The facilitation capacity-building aspect was challenging to explain, however, and created confusion - until one of the student governors realised that we wanted them to 'be us' - i.e. take on our own roles. In retrospect, it might have been helpful to work through all the activities once first, before separately focusing on the elements of good facilitation and allowing the students to practice facilitating each other.

5.4.4. Identification and closure of value-action gaps: some preliminary reflections

The toolkit activities were well received by the student governors during the capacity building sessions, and generated some lively, and generally positive, discussions. In the first session, the collective story of 'the future we want' generated a number of new SRL skill statements that were not present in either the original WeValue list or the education reference list. These included, among others:

- Evaluating what's important to us and what isn't
- Looking after ourselves and our families
- Not being so dependent on technology that we lose the ability to write and socialise (communicating face to face; spending quality time with people)
- Accepting others instead of judging them
- Choosing jobs that we love instead of only thinking about how much we can earn

We noted that this exercise, by its very nature, inherently required participants to reflect on value-action gaps. Thinking about the future that they would most like to see for their school, in an ideal world, helped to focus their attention on things that matter to them but may not be fully enacted in the school at present. We also noticed that while the 'voting' with chocolates was a popular activity, it was the grouping exercise in the second session that appeared to stimulate the deepest reflection. It led to some important realisations about how different SRL skills are interconnected, and a revaluing of some statements that had initially been seen as unimportant.

The `human survey' / spatial survey seems to be another helpful tool for assessing values enactment and highlighting value-action gaps, and an important observation was that consensus among students is not necessary in order for the exercise to be useful. On one of the three chosen skills, "Maintain a sustainable society, e.g. recycling, energy", there was a strong consensus that the school was not doing enough and that these issues should be taken more seriously by the senior management. On the others, however ("Be less judgemental – accept people more" and "Not to become so dependent on technology that we lose the ability to write and socialise"), there was a wide spread of responses –ranging, in the latter case, from around 10% to 80% agreement. This prompted lively discussions, which resulted in some students changing their positions in the 'survey'.

Finally, the non-cognitive approach of role-play proved very powerful, enabling students to embody the two contrasting situations of judgement and acceptance. (Due to time constraints, only one skill was role-played). Perhaps understandably, students spent more time enacting well-known problems than envisioning workable solutions, and we realised that the activity guidelines could be reworded to encourage future facilitators to focus on the positive. Nonetheless, participants understood the point of the exercise and contributed meaningfully to a follow-up discussion about what could be done differently. While some suggested that the senior management should take a tougher stance on bullying, others acknowledged that they themselves - as peer leaders - could play a role in helping to create a climate where everyone feels accepted and valued.

We envisage that these processes of identifying and closing value-action gaps could be taken further, e.g. by asking students to reflect on their chosen SRL skills through arts-based activities (painting, poetry, music, dance, monologue, etc.) and then to identify specific, measurable actions that they can take themselves and/or request the senior management team to implement. The senior management, in return, might pledge to implement a minimum number (e.g. three) of the viable suggestions made by the youth for building a better future at the school. These new activities have not yet been tested at the time of writing.

5.5. Discussion

Although this project is still ongoing, it has already demonstrated its utility at several levels. First, we have shown that each of the 'problematic situations' outlined in Table 5-3 is beginning to shift towards its respective 'preferred situation' - albeit to a limited extent, in the light of resource constraints. We have demonstrated that it is possible, on a small scale and with an amenable group of students, to (1) adopt a positive and constructive approach to the teaching of EfSRL, which focuses on developing values and skills necessary for envisioning and co-creating better futures; (2) identify value-action gaps, and at least begin to understand how they might be closed; and (3) design a toolkit of values-based indicators suitable for a school context. It is important to note, however, that we were working with peer-elected student governors, who might be more engaged, positive and proactive than the general population in their age group.

We have not yet established whether students aged 14-16 can be trained to work effectively as facilitators for a younger age group, as this has not yet been carried out due to examination schedules. In addition, we have not yet explored the full potential of the toolkit activities for closing, rather than merely identifying, value-action gaps. We anticipate, however, that arts-based reflection may be valuable for helping students to identify specific action points - both for themselves and for the senior management team. In this respect, the

willingness of senior management to listen to students and implement their viable suggestions is crucial, as it could be profoundly empowering for the youth to see their work leading to observable changes within the school. (The 'school evaluation toolkit', still in construction, could potentially be useful at this point). The potential role of class teachers also needs closer attention, as in our work with student governors the teaching staff were only minimally involved, although we have since demonstrated some of the activities to a Year 8 PSHE teacher and his students within a classroom context,

5.5.1. How does this work contribute to values literature?

Relating our findings to literature, we suggest that our work links Maio's (2001) hypothesis - that the attempt to articulate ethical or pro-social values in words can reinforce and strengthen those values, where they might otherwise be 'over-ruled' by more urgent needs such as money, comfort or time – to the field of EfSRL. We have extended the nascent literature on values-focused evaluation (Burford, Hoover, et al., 2013; Burford, Velasco, et al., 2013) by illustrating that in principle this approach can be modified for formal education settings. However, congruent with the work of Flores (2008) on Youth Participatory Evaluation, we suggest that it is also important to elicit values statements from young people in their own words, as a process based on analysis of what is important to teachers may not capture everything that matters to students. The combination of an explicit values elicitation step with reflection on a pre-existing 'menu' can ensure that participants are both empowered to express whatever is already important to them, and challenged with new ideas that they might not previously have thought about.

An observation made during the first capacity building session with student governors has important implications for the conceptualisation of values, and may point to a second, hitherto unreported, type of 'gap'. We learned that the students felt they were already practising many of the skills described in the list, often outside the core curriculum; but they had neither articulated them in words, nor previously thought of them as `skills for sustainable and responsible living'. This observation echoes a statement by Rescher (1982) that value subscription can manifest itself both through discourse (what people say) and

through overt action (what they do), but the critical test of value presence is consistency between the two. Citing Rescher's work, Schlater and Sontag (1994, p. 5) offer two contrasting examples of inconsistency: "A person may 'talk' the value but not implement it in action, or a person may act in accordance with a value but not subscribe to it verbally."

By analogy with value-action gaps, the second inconsistency described by Schlater and Sontag (1994), i.e. a situation in which people are known or hypothesized to hold certain values but do not talk about them, might be termed a value-discourse gap. This can be related to comments by teachers that several of the skills in the reference list were barely covered in (or even, in a few cases, were entirely absent from) the UK national secondary curriculum. Merely by introducing them as topics of conversation, and linking them explicitly to SRL, the toolkit has already contributed towards the closure of value-discourse gaps.

5.5.2. How does this work contribute to design literature?

While still in its early stages, this work underscores the importance of involving all relevant stakeholders in sustainable design processes (Blizzard & Klotz, 2012) and in particular, highlights the need for meaningful involvement of youth.

While there is a vast literature on different aspects of co-design and participatory design, very little of this work refers directly to the participation of children and young people, with some notable exceptions in the fields of architecture (Driskell, 2002; Spencer & Blades, 2006) and information systems (Druin, 2005). It can be assumed that most 'participatory' design – even in school contexts – remains dominated by adults, with the participation of children and youth primarily at a tokenistic level (c.f. Hart 1992). This is analogous to the situation in the field of evaluation prior to the seminal work of Kim Sabo Flores (2008; see also Hochachka, 2005; Seamon and Zajonc 1998), and it can therefore be assumed that there may be a productive crossover between Youth Participatory Evaluation and Participatory Design – not only in terms of specific methods and strategies, but also underlying assumptions

about youth and their capacities. We will explore these ideas in greater depth in our future work with the toolkit.

5.6. Conclusions

In this chapter, we have presented preliminary findings from an ongoing project aimed at creating a values-based EfSRL toolkit for secondary schools. We have illustrated that, even at this early stage, the project has achieved some progress towards creating artefacts that can transform problematic social situations into preferred future situations.   The framing of the designed artefact – the toolkit text – embeds within it an emergent body of knowledge about what is valued by teachers in specific settings in the UK, Ireland and Tanzania, and (to a limited extent) by 14-16 year old students attending a secondary school in southern England. In addition, through the nature of the designed artefact and users' experiences with it, this study has contributed to the development of theory about values: confirming the applicability of earlier work on value-action gaps to the new context of schools, and identifying a new type of `gap' – the value-discourse gap – that has previously been overlooked.   Finally, within the field of design research, we have drawn attention to the importance of participation by children and youth and highlighted a potentially useful overlap with Youth Participatory Evaluation.

Chapter 6:

Discussion and conclusions

Chapter 6. Discussion and conclusions

6.1. Aims and approach of this chapter

In this chapter, I critically review the work presented in the four published papers, with a focus on (a) disrupting and challenging the underlying assumptions that I made, in accordance with my personal experiences and perspectives, at the time when the research was undertaken; (b) expanding on areas that were not sufficiently theorised in the publications; (c) reflecting on the enduring significance and wider implications of the papers, especially in the context of the shifts in the global political landscape since the first paper was published in 2013; and (d) correcting some minor omissions and oversights that were missed by the peer reviewers.

I have chosen to do this on a sequential basis, one chapter at a time – which inevitably involves a certain degree of repetition, but has the advantage of ensuring that the links to the original work are not lost.   The final section brings together insights from all four papers to lay the foundations for an emerging research agenda, which would continue the theoretical work described in this thesis and embed it in practical applications in the fields of education, project evaluation, and community engagement.

6.2. Reflections on Chapter 2 (Organisational Impacts of Values-Based Evaluation)

This paper explores the organisational impacts of using 'values-based' approaches to project evaluation and performance assessment in non-formal education. It provides an empirical basis for the claim that inductive and intersubjective approaches to evaluation design might be useful and beneficial in practical contexts, illustrating that small-scale field trials with values-based indicators and assessment tools had a variety of positive impacts. However, one of the key omissions from the original paper was a discussion of limitations and caveats, relating both to the study itself and to the general approach. Some reflections on these are as follows.

6.2.1. Limitations of the research study

As acknowledged in the paper, this was an exploratory study carried out with a small sample of organisations. In addition, the nature of the research method – a re-analysis of an existing dataset, consisting of project reports, interviews with managers of the respective organisations, and interviews with the researchers who carried out the field visits – resulted in one-off 'snapshots' of impact. If the data collection process had been designed from scratch to answer research questions about organisational impact, it would have included more systematic follow-up of each organisation over a period of weeks or months. This would have allowed the research team to state, with a higher degree of confidence, that the reported outcomes were (a) noticed by multiple team members, rather than relying on the testimony of a single individual, (b) long-lasting, and (c) attributable to the evaluative interventions rather than to some other factor.

Another important limitation that was not highlighted in the paper is that only two of the eight case study organisations were private sector companies – the remainder were civil society organisations (CSOs) or university departments that already had an environmental and/or social sustainability focus. This meant that the 'proto-indicators' or 'trigger statements' that were presented as catalysts for reflection – which were co-designed by other CSOs promoting education for sustainability - were inherently more likely to resonate with staff in the target organisations than in private sector companies with no direct interest in sustainability concerns. The generalisability of the research beyond the arena of education for sustainability may therefore be lower than I initially suggested. However, this does not affect validity *within* this arena, and indeed may be seen as enhancing it, in the sense of providing additional evidence that these approaches can be useful.

This study does not provide any substantive evidence that the WeValue approach can 'contribute to the development of ethical cultures', as I had provisionally claimed (see section 2.4.4). Rather, what the results do show is that, as I had stated in Section 2.4.1, people are enabled to 'articulate in discourse what they already undertake or experience together as values-in-action'. In other words, intersubjectively generated proto-indicators can be useful for helping managers to verbalise shared values more clearly – and in particular, to mainstream them more explicitly into areas such as internal and external communications,

training, strategic planning and performance assessment – in organisations that *already have* strong ethical cultures, rather than necessarily contributing to their development in organisations where they did not formerly exist.

Comparison of this paper with Chapter 5, which was drafted around two years later (even though the actual publication date is earlier) and thus represents an evolution of my thinking, highlights a crucial point in respect of this issue. The concept of 'value-action gaps', also called 'attitude-action gaps' or 'values/behaviour gaps' (people's tendency to act in ways that are incongruent with the conversations they have, or surveys they complete, about their values) is well established in literature (Bardi, Lee, Hofmann-Towfigh, & Soutar, 2009; Ben & Potter, 2007; Blake, 1999; Kennedy, Beckley, McFarlane, & Nadeau, 2009; Kollmuss & Agyeman, 2002). However, in Chapter 5 I have identified a previously undescribed converse phenomenon, which I have termed 'value-discourse gaps': namely people's tendency to be unaware of, or reluctant to talk about, values that they are already enacting.

When Chapter 2 is revisited in the light of this key insight, it becomes apparent that inductively- and intersubjectively-designed evaluations and assessments can provide an extremely effective means of contributing to the development of ethical *conversations* – in other words, bridging value-discourse gaps. Whether or not these conversations are ever translated into actual changes in *culture* within the respective organisations (i.e. bridging value-*action* gaps) will depend on a number of complex and intertwined factors, including the organisational structure (hierarchical, flat, or somewhere in between); the prevailing power relations, especially in relation to dominant versus marginalised groups; and who is included in, and excluded from, the inductive and intersubjective processes. Nonetheless, the very existence of the conversations is noteworthy, and could provide a starting point for several avenues of new research.

6.2.2. Limitations of the approach

Following on from the point above, an important caveat relating to the specific evaluation approach described in this paper is the risk that it may exacerbate existing hierarchies, bureaucracies and problematic power dynamics, especially in organisations that do not have a strong Equality and Diversity ethos.

The 'WeValue' process for developing intersubjective values-based indicators, whether for evaluation or only for the purpose of shared reflection, is based on the assumption that every individual participant can and will openly and honestly articulate what they find 'valuable, meaningful and worthwhile' (in relation to the specified practical context) during the elicitation phase. This is usually achieved through verbal contributions, or directly attributable written contributions, within a workshop context. These, along with any additional prompts or trigger statements that may be supplied by the facilitator, are then placed on the table for discussion during the challenge phase.

It has been established in the process of developing indicators and facilitator training programs, but not specifically articulated in this paper, that the 'WeValue' workshop facilitators are strongly encouraged to elicit contributions from people of lower status within the organisational hierarchy (e.g. service users, interns, volunteers or junior members of staff) before allowing contributions from the senior managers. However, this assumes that a person's position within a formal hierarchy is the only relevant dimension of power. It obscures the difficulties that may be experienced by members of marginalised groups, such as children and youth, the elderly, single parents, people with disabilities or chronic illnesses, neurodiverse people, Black, Asian and Minority Ethnic (BAME) people / people of colour (PoC) in majority white communities, people belonging to religious minorities, Indigenous peoples in countries that have been colonised, people from less privileged socio-economic groups, women in patriarchal societies, or people who are lesbian, gay, bisexual, transgender, queer, intersex or asexual (LGBTQIA+) in cis-heteronormative contexts – irrespective of their position within formal hierarchies.   As highlighted by Kimberlé Crenshaw's intersectionality theory (Crenshaw, 1989), these identities are not experienced independently, but can interact to leave certain people (such as trans women of colour) doubly or multiply marginalised by intersecting systems of structural, political, and/or representational oppression.

The first crucial question is whether members of marginalised groups, and especially those experiencing double or multiple disadvantage, are represented in the workshops at all. Some aspects of marginalisation, such as age and chronic illness, would statistically reduce the likelihood of being present in the workplace for any type of workshop; while for others, such as race or sex, either overt discrimination or unconscious bias might reduce an individual's chance of being selected. In the research, I did not address questions such as how

the 'WeValue' workshops were convened, which processes and criteria were used to select participants, and whether the selection processes might have been influenced by prejudices against certain groups.

Assuming that they are indeed present at the table, there may be several reasons why members of marginalised groups are unable or unwilling to talk openly and honestly about what they find valuable, meaningful or worthwhile. Conversations about values are particularly susceptible to social desirability response bias (Arnold & Feldman, 1981; Fisher & Katz, 2000; Furnham, 1986), especially when there is a strong power differential. In addition to social desirability response bias and group conformity bias, which are often subconscious, individuals may consciously avoid articulating a position that runs counter to the mainstream because they fear losing their status or acceptance within the group - or even their jobs. In some cases, particularly LGBTQIA+ and Indigenous people, individuals' status as members of a marginalised group may not be known to everyone before the workshop; 'outing themselves', whether directly by stating their affiliation or indirectly by affirming certain values, could place them at increased disadvantage and may even be dangerous. In other cases, such as people with autistic spectrum conditions, attention deficit disorders, hearing impairments or verbal communication difficulties, the act of contributing opinions to a workshop-type discussion could be problematic in itself.

Even if potentially challenging or controversial viewpoints are articulated during the first phase of the workshop – namely, the 'elicitation' phase - this does not inevitably mean that these viewpoints will be incorporated into the workshop outputs. This is because the second phase of the 'WeValue' process (the 'challenge' phase) typically involves individual or group reflection on a list of proto-indicators or trigger statements, which has been created to prime specific values. The indicators and/or assessment tools that constitute the final workshop outputs typically include a mixture of statements from the first phase of the workshop (directly elicited by asking participants about what they find valuable, meaningful or worthwhile) and those from the second phase (primed with the list).

In this way, the diversity and motivations of the group responsible for constructing the proto-indicator 'menu' are key factors in determining the extent to which the voices of marginalised participants are either strengthened or erased. So, too, are the background, training and prior experiences of the facilitator(s), especially in relation to their own privilege

and the extent to which they are aware of it. Without a concerted attempt to ensure that the menu reflects the full diversity of the organisation, the default situation is that at least some people will be erased. However, with sufficient investment of financial and human resources, materials for the challenge phase could be created with the explicit aim of 'decolonising minds' and pushing people, especially those with privilege, to think outside the comfort zones created by their own societal norms. This constitutes a clear example of transformative and transgressive learning, which would require considerable attention to facilitator training in order for it to be successfully achieved.

The original ESDinds project that generated the WeValue process was led by a consortium of four civil society organisations, two university research groups and an independent advisor (ESDinds, 2011). While gender balance was actively promoted, two of the members of the steering group were Latin American and the others were white European: there was no representation from Africa, Asia, Australasia or any Native / Indigenous group in the project. Another important point is that at least five of the ten members of the original steering group (including the team leader and the independent advisor) were followers of the Baha'i faith, and had met or connected with each other initially within that context. Of the four civil society organisations that were official project partners, two described themselves as Baha'i-inspired, and a third was an inter-religious body (the Alliance of Religions and Conservation). An additional Baha'i organisation was originally listed as a formal partner, but later dropped out of the project. Therefore, although the Baha'i 'roots' of ESDinds were not explicitly acknowledged during the project or in any of the reports or publications arising from it, it is probable that subconscious biases in favour of Baha'i values and principles influenced the direction of the research, including the original choice of value-labels ('Unity in Diversity', in particular, is a key phrase that is widely used within the Baha'i movement: see, for example, Baha'i International Community, 1998) and the wording of the proto-indicator 'menu' that was used in the field studies.

The potential for proto-indicator menus, or other tools used during the 'values challenge' phase, to be used to covertly promote a particular ideological agenda has already been raised in Chapter 3, and will be discussed in more detail below. This specific example of Baha'i influence is perhaps not directly problematic in terms of what has been *included* in the proto-indicator menus, in the sense that most of these 'Baha'i-inspired' values and principles

are accepted tenets of education for sustainability. They include individual responsibility, respect for ethnic and cultural diversity, gender equality, concern with social justice and the empowerment of marginalised people, and the elimination of extremes of wealth and poverty (Baha'i International Community, 1998, 2002). The menus resulted from extensive discussions and successive rounds of prioritisation, which included numerous representatives of secular organisations as well as Baha'i-inspired ones, and there are no proto-indicators that attempt to proselytise. However, the prospect of a hidden religious bias does raise concerns about which values and perspectives may have been *excluded* from the menus, by virtue of not being considered sufficiently important by the groups responsible for developing the menus. The Baha'i position on sexuality, for example, is that all sexual activity is morally wrong unless it is within the context of marriage between a man and a woman, and that same-sex attraction is curable through medical intervention and prayer, although Baha'is are urged not to discriminate against people with a 'homosexual orientation' (Research Department of the Universal House of Justice, 1993).

It is interesting to note, with hindsight, that the initial iteration of the proto-indicator menu (used in most of the field trials) did not include any mention of non-discrimination on the grounds of sexual orientation. Its presence in a later iteration (as a passing mention in one proto-indicator, alongside other aspects of diversity) was directly attributable to my own editing work. The fact that 'sexual orientation' was mentioned explicitly in the menu, whereas 'gender identity' was not (and presumably still is not, to this day) clearly reflects the fact that in the period between 2011 and 2016 I was extensively engaged in questioning my own sexuality, but as yet unaware of conversations around transgender identities and challenging the validity of the gender binary. This provides an illustrative example of the sensitivity of the menu design process to individual biases, prejudices and blind spots.

Another concern about the approach, which I have already mentioned in the paper, is that 'findings from previous studies might be negated in cases where individuals do not have shared values-in-action to build on but are trying to reconcile superficial values with no grounding in a context-specific reality' (Section 2.4.4). I have suggested that in such circumstances, values-based approaches could be counterproductive rather than useful, potentially leading to misunderstandings and polarization. However, in the paper I did not develop this argument to its logical conclusion, which is that there are certain situations in

which it may be inappropriate to use the 'WeValue' approach to develop indicators and assessment tools. These include, but are not necessarily limited to, new organizations; groups of people who do not usually work together, even within the same organization; and short-term or recently-launched projects.

6.2.3. Wider implications

In the introduction to this thesis, I introduced the 'WeValue' approach not only as a tool in itself, but as one named example of an inductive / intersubjective process for designing indicators that reflect what people collectively find valuable, meaningful or worthwhile. Some of the concerns expressed above in relation to the 'WeValue' approach are similarly applicable across the full spectrum of inductive / intersubjective indicator design approaches, regardless of the specific context. One is that indicators of any sort (along with the wider development apparatus, described by Ferguson (Ferguson, 1994)as an 'anti-politics machine') can serve as a mechanism for obscuring political issues and/or reframing them as technical problems. When indicators are developed inductively/intersubjectively and are concerned with intangible qualities such as values, mindsets, motivations and intentions, the potential for both depoliticization and the consequent expansion of bureaucratic power becomes even greater. Thus, the risk that intersubjective indicator design processes will reinforce existing hierarchies and power differentials - silencing the marginalised, while amplifying the voices of those who already hold power - is not contingent on the type of design process. Likewise, the requirement for a meaningful context of shared practical experience among those engaged in the indicator design process (as one of the key criteria for the creation of shared understanding through intersubjectivity) still applies across the board, regardless of the specific approach that is used to generate the indicators.

6.3. Reflections on Chapter 3 (Operationalisng the 'Missing Pillar' of Sustainability)

In this section I will discuss two issues, raised by reviewers of my original submission to the journal *Sustainability*, that I addressed briefly in the final published version of the paper but still did not fully explore. These are (i) the extent to which Nicolescuian transdisciplinarity provides a theoretical lens through which intersubjective indicator design can be more fully

understood; and (ii) the potential for perverse effects resulting from inadvertent misuse, or deliberate abuse, of intersubjective approaches to indicator design.

6.3.1. Theorising 'values-based indicators' through a transdisciplinarity lens

During peer-review, I was asked to comment on the transdisciplinary nature of my research, with specific reference to the work of the Romanian theoretical physicist Basarab Nicolescu. In the revised paper I made only a passing reference to Nicolescuian transdisciplinarity, and the crucial importance of transdisciplinary approaches to research and education in transitions to sustainability. While mentioning Nicolescu's 'three axioms of transdisciplinarity' as interesting and potentially relevant, I neither explained them in detail, nor adequately acknowledged them as a fundamental theoretical basis underpinning intersubjective indicator design. This section seeks to remedy that omission.

(a) **The ontological axiom** states that there are 'different levels of Reality of the Object and, correspondingly, different levels of Reality of the Subject' (Nicolescu, 1998, 2002, 2006, 2010). These multiple levels of Reality have been summarised by McGregor (2012) as respectively the *external world* where information flows, comprising environmental, economic, and cosmic/planetary realities, and the *internal world* where consciousness flows, comprising individual, political, social and historical realities. They are brought together by the *hidden third,* or 'zone of non-resistance', which mediates the flow of consciousness and information and is expressed through culture, art, religions and spiritualities. The zone of non-resistance, as noted by Nicolescu (2010, p. 26), 'corresponds to the sacred – to that which does not submit to any rationalization'. The external world, the internal world and the 'hidden third' can be correlated, respectively, with the objective, subjective and intersubjective domains described in my paper.

(b) **The logical axiom** states that 'passage from one level of Reality to another is ensured by the logic of the included middle' (Nicolescu 2010, p. 24). This contradicts the logic of the excluded middle, characteristic of classical physics, which states that it is impossible for an object to be both A and non-A at the same time. Applying the logic

of the excluded middle within the social, cultural, religious and political spheres creates dichotomies - 'good/evil', 'heaven/hell', 'women/men', 'rich/poor', 'white/black', and many more – which, in turn, almost invariably lead to harmful ideologies such as racism, xenophobia, religious fundamentalism and antisemitism. Other examples, not explicitly mentioned by Nicolescu, are sexism, homophobia and transphobia. The logic of the included middle, which is at the heart of quantum physics, relies on 'a third term T which is at the same time A and non-A' (ibid, p. 30) and is situated at a different level of Reality from the apparent opposites.

(c) **The complexity axiom** states that 'the structure of the totality of levels of Reality or perception is a complex structure: every level is what it is because all the levels exist at the same time' (Nicolescu, 2010, p. 24). What this means is that everything is connected, or as Nicolescu explains it, 'complexity is a modern form of the very ancient principle of universal interdependence' (ibid, p. 31).

The way in which these axioms underpin inductive and intersubjective indicator design is as follows. Firstly, sustainability indicators have traditionally focused on the 'external world' – the collection of environmental and economic data – and been assumed to be objective in nature, while discussions of values have been confined to the 'internal world' expressed through psychology, politics, sociology and history. However, the creation of inductive and intersubjective (or 'values-based') indicators within the realm of Education for Sustainability requires practitioners to transcend this apparent dichotomy and open their minds to uncomfortable new perspectives. On the one hand, indicator specialists who are accustomed to scientific measurement are forced to confront issues such as emotions and subjectivity. On the other, educators who are already working with values, mindsets and motivations - usually in rather vague and intangible ways - must wrestle with the challenges of measurability and quantification. In order for this to be achieved successfully, a 'zone of non-resistance' (the 'hidden third' or 'included middle') must be created. This is a space in which people can connect with each other as human beings, overcome their prejudices and recognise their interdependence, before proceeding to work towards a shared transdisciplinary understanding (a process that is necessarily transformative and frequently transgressive). The emergence of new insights, intelligence and knowledge through the

exchange of diverse stories and experiences has been termed 'intellectual fusion' (McGregor & Volckmann, 2013).

The 'WeValue' approach to indicator design developed through the EU-funded ESDinds project (cited in this thesis as a worked example, rather than the totality, of inductive / intersubjective indicator design) was neither explicitly grounded in Nicolescuian transdisciplinarity, nor directly informed by the transdisciplinary research literature. Nonetheless, the steps that were undertaken to create the zone of non-resistance in the project closely resembled those that might have been expected if it had been undertaken from a Nicolescuian standpoint from the start:

- ***Culture and art:*** as the initial step in conversations about values, people were invited to share brief narratives about their past experiences and/or their visions for the future in relation to the specific project or organisational context that they wanted to evaluate. Sometimes, photographs were used as a visual prompt for the storytelling exercise, and on other occasions, people were invited to make a sketch or doodle as a precursor to sharing their insights and experiences.

- ***Spiritualities****:* the content of these 'values elicitation' conversations was focused on understanding what people found most ''valuable, meaningful or worthwhile' within their project or organisational context. This meant that although it was not described as such, the discussion of values was always rooted in conversations about meaning and purpose - core components of secular spirituality.

- ***Religions****:* as discussed above, there may have been a hidden bias in favour of principles and values espoused by members of the Baha'i faith in the original ESDinds project.

Thus, the ESDinds project – however inadvertently – provided the ingredients for the generation of transdisciplinary knowledge, as outlined in Nicolescu's model of 'levels of Reality'. Through the application of art and culture (storytelling, photography and visual art), spiritualities (discourses of value and meaningfulness) and perhaps some degree of latent or tacit religious influence, a zone of non-resistance was created. This permitted a productive dialogue between the 'internal world' of personal experiences, social norms, organisational politics and historical contexts, on the one hand, and the 'external world' of observable project outcomes and measurable indicators, on the other. Thus, the

central theoretical argument of my work in this paper, namely that 'values can be measured when they are intersubjectively conceptualised within a clearly specified practical context', is fundamentally a transdisciplinary argument. The process of discussion that generates intersubjective indicators is the 'included middle': the level of Reality at which 'subjective' values and 'objective' data collection efforts can be comprehended simultaneously and are no longer viewed as opposites.

6.3.2. Value conflicts and the potential for perverse effects

In responding to a reviewer's question about the potential for perverse effects to arise from the use of values-based indicators, I acknowledged this as a possibility and suggested ways in which it might be mitigated. At the time of writing this paper in 2013, however, I had not fully appreciated the extent to which the creation of a predetermined 'menu' of prompts or reference indicators might influence the user(s) to adopt the values held by the designer(s) of these materials – and, consequently, the extent to which values-based indicators could potentially be transformed from a tool for empowerment to a mechanism of repression. The emergence of this as a significant concern is directly related to changes in the global political landscape since the paper was first published.

Coming to the research from a Baha'i perspective in 2010, having recently volunteered as a tutor for Baha'i study groups in Tanzania, I based the original work on the assumption that values such as 'unity in diversity', 'respect and care for the community of life', 'justice', 'trustworthiness' and 'empowerment' were held as near-universal ideals within the United Nations and among the majority of its Member States. When this paper was published in 2013, I still saw it as reasonable to assume that the 2030 Sustainable Development Goals and their indicators would be created, implemented and evaluated in such ways as to promote these and similar values. However, recent developments in international politics since the summer of 2016 have served to highlight the fact that many political decisions (with far-reaching implications) are rooted in very different kinds of values and motivations. They include, among others, the Brexit movement in the UK, the election of Donald Trump as President of the United States, and the global rise of 'alt-right', neo-fascist and nationalist movements. What appeared in 2000 as a clear global

trend towards decolonization, cultures of collaboration and an ethical framework for sustainability (as exemplified, for example, by the UN Millennium Declaration and the Earth Charter) has now shifted towards a combination of growing nationalism and corporate protectionism – in effect, a recolonization. The implications for the Sustainable Development Goals are yet to be seen, but Trump's 2017 decision to withdraw the United States from the Paris Agreement on Climate Change gives a preliminary indication of what may lie ahead.

In a 'post-factual' society, in which politicians are at liberty to dismiss or even ridicule scientific evidence that does not support their ideological position, the use of intersubjective indicator design at a national level is not merely problematic but potentially dangerous. it is easy to imagine that 'values-based' approaches to project evaluation and organisational performance assessment could potentially be co-opted to further a state-sponsored agenda of repressing minorities and political dissidents, while neglecting or overriding genuine environmental and social justice concerns. This will be explored further in the discussion of Chapter 4.

6.4. Reflections on Chapter 4 (Indicator Design for Complex SDGs)

6.4.1. Limitations and critiques

This paper, the most recently-written of the four, explicitly acknowledges the broader concept of inductive indicator design beyond the specific case of the 'WeValue' approach. It raises, but does not unpack, the suggestion that inductive / intersubjective processes for designing indicators can have 'important implications for engagement in and the democratisation and diversification of global sustainability and development agendas as a whole' (Section 4.4.8) and might 'open up the possibility of creating a place at the SDG table for other sustainability discourses that do not sit comfortably with the implicitly neoliberal concept of sustainable *development'* (c.f. Robinson, 2004).

Among these discourses, one, namely David Abram's seminal discussion of the multiple subjectivities of the 'more-than-human' community of life (Abram, 1996), is

particularly noteworthy. This is on account of the stark contrast between its ubiquity in Indigenous epistemologies and shamanic practices, on the one hand, and its outright neglect by the international community in conversations about 'ESD' and 'education for sustainability', such as the UNESCO Decade for ESD or Target 4.7 of the SDGs, on the other. Abram's overarching argument is that it is crucial for academics to stop viewing 'the environment' merely as a component of the objective domain – a generic 'it' that can be treated as a source of natural resources for human activity or, at best, as a locus for human healing– and, instead, 'enter into relation with other species on their own terms'. This would entail a renunciation of the 'perceptual boundaries demarcated by [our] particular culture… reinforced by social customs, taboos, and most importantly the common speech or language' (ibid, p. 9).

If such perceptual boundaries existed in 1996, fuelled (as Abrams argues) primarily by the western fixation with the written word, it would seem likely that in the intervening decades they have been substantially reinforced by the rise of social media and other electronic communications.   There is a certain irony in the observation that even the experience of walking the land is increasingly mediated by technology, via mobile phone apps that urge people to download well-established routes or upload photos of their own favourite walks (e.g. www.gojauntly.com, www.ifootpath.com) – simultaneously encouraging an increased exposure to outdoor environments, and a further disengagement with the more-than-human world as a community of perceiving subjects.

I have raised the possibility of a deeper engagement with intersubjectivity in the sense of acknowledging and valuing multiple subjectivities within the community of life, along the lines proposed by Abram, in a working paper entitled *Collaborative Research for Sustainability: An Inside-Out Design Manifesto* (Burford, 2015). This paper - aimed primarily at interdisciplinary and transdisciplinary researchers in the arts and social sciences, rather than specialists in sustainability -was written in response to a call issued by Keele University's Community Animation and Social Innovation Centre (CASIC) for papers that engaged with the subject of collaborative or co-produced research in intriguing or novel ways. Accordingly, it incorporates visual art and poetry alongside the more traditional academic prose.   The purpose of this is twofold: firstly, to provoke an affective response alongside the cognitive one in the hope of achieving a more immediate and longer-lasting impact, and secondly, to

disrupt and challenge the norms of academic writing from the standpoint of an artist as an 'entrepreneur in conventions' (Kagan, 2011). As I explain it in the paper (Burford 2015, p. 6):

> *'My position on these issues draws on the seminal work of David Abrams [sic] (1996) in 'The Spell of the Sensuous: Perception and Language in a More-than-Human World', and as such, I will not expound it in depth. In subverting the conventional practice of 'peer review' in Article 3, my aim is to use humour to bring Abrams' [sic] arguments out of the comfortable niche of ecophilosophy / ecopsychology (where they have resided for nearly two decades already) and into the mainstream: inviting academics to reflect on the wider implications of their work for the more-than-human community, and to attempt an empathy with other-than-human perspectives.'*

The 'Article 3' referred to in the text is a section of my 'manifesto', which takes the form of a long poem. The following is the excerpt in question (Burford 2015: 10):

> *'listening to the forest should be the first lesson,*
>
> *a courageous and radical act*
>
> *that flies in the face of convention*
>
> *we listen to the robin's song*
>
> *and the rustle of leaves, not as detached*
>
> *observers of objects, but guests observed*
>
> *by hosts in their homes: hearing their peer reviews*
>
> *of our well-meant efforts, their plaintive questions*
>
> *demanding responses, calling us (as so-called experts*
>
> *who thought our truths self-evident) to revise and resubmit…'*

The manifesto both exemplifies the translation of the work presented in this thesis for a specific audience, and provides an illustrative example of how it might be applied in a new context.

The conversation around acknowledging multiple subjectivities within the 'more-than-human' world is not, of course, the only aspect of Indigenous philosophy that is missing from the model of sustainability competencies set out in Paper 4. Indeed, the model could be heavily criticised for being Eurocentric in its essence, and thereby contributing to the ongoing marginalisation of Indigenous perspectives. Besides Abram's multiple subjectivities, another key omission is the 'Sacred World' or spiritual dimension of existence, which is not acknowledged in any of the 26 competencies in the UNESCO model or the three additional ones identified by Wiek et al. (2011). As described by Dennis Foley (2003), Indigenous philosophy within an Australian context is based on the triangulation of the Physical World, Human World and Sacred World, which Foley defines as follows (pp. 46-47):

> '*The physical world is the base that is land, the creation. The land is the mother, and we are of the land. We do not own the land, the land owns us. The land is our food, our culture, our spirit and our identity. The physical world encapsulates the land, the sky and all living organisms. The human world involves the knowledge, approaches to people, family, rules of behaviour, ceremonies, and their capacity to change. The sacred world is not based entirely in the metaphysical, as some would believe. Its foundation is in healing (both the spiritual and physical well being of all creatures), the lore (the retention and re-enforcement of oral traditions), care of country, the laws and their maintenance.*'

Thus, a broader critique of Paper 4 is that in attempting to argue a case for inductive and intersubjective sustainability indicators, I have constructed the conceptual framework against which these indicators are assessed – the set of 'skills needed to promote sustainable development' – in a way that excludes Indigenous and other non-mainstream perspectives. The blind spots that affected the construction of the assessment framework are very similar to those that have limited the development of the indicators themselves. The wider implications of this will be discussed below, in the context of defining a research agenda for the future.

6.4.2. Shifting the target audience: from 'global' to `glocal'

An important concern that I have hinted at but not sufficiently emphasised in Chapter 4 is that the process of inductive / intersubjective indicator design does not come with any inherent safeguards or limitations, in terms of the values motivating its use or the purposes for which it can be applied. Like nuclear technology, it could be developed for purposes that are broadly peaceful and positive, and subsequently co-opted to serve the interests of those with power and privilege. In this context, it could be viewed as highly irresponsible to promote the application of inductive / intersubjective indicator design (such as the creation and use of values-based indicators) in bureaucratic and 'top-down' contexts, whether at national, regional or global levels, for the reasons that I have discussed in section 6.2 above.

Instead of advocating a change in the official mechanisms by which the Sustainable Development Goals are operationalised as measurable indicators (which in retrospect was always unrealistic because of the timing of the publication, the limited reach of the journal, and the rigidity that often characterises bureaucratic processes) it might be more advisable to focus on initiatives that may be described as 'glocal', i.e. arising from the convergence of global dynamics and local contexts (Swyngedouw, 2004). These could include, among others, transnational networks of local organisations that are sharing experiences and working towards common goals; projects bringing together researchers and practitioners from different countries as co-equal partners, especially those with an explicit focus on transdisciplinarity and/or sustainability, such as the 'T-learning' network on transformative and transgressive learning, the European Transdisciplinary Design Network, and the td-net network for transdisciplinary research. It could also be of benefit to programs such as the Earth Charter Initiative that use core documents or toolkits as catalysts, and unifying principles, for a multiplicity of small-scale individual and group initiatives around the world.

6.5. Reflections on Chapter 5 (Promoting Sustainability Skills in Schools)

6.5.1. Contrast with earlier papers: focus on reflection rather than evaluation

This paper provides a bridge between the vast literature on education for sustainable development (ESD), environmental education (EE) and education for sustainable and responsible living (EfSRL), on the one hand, and participatory evaluation, on the other. It does this by highlighting the fact that indicators may have two separate roles within an organisation: as catalysts for reflectionthat can stimulate teaching and learning, and/or as starting points for evaluation and assessment that can promote organisational development. The former role is an example of what Gudmundsson (2003, p. 5) describes as the `conceptual use' of indicators, opening up new spaces for conversation and shifting people's perceptions of what constitutes a valuable outcome; the second exemplifies the 'instrumental use' of indicators for the collection of data.

In contrast to Papers 1 and 2, which both focus on applications of intersubjectively-designed indicators (specifically values-based indicators) to evaluation, this paper is primarily concerned with the use of indicators as catalysts for reflection within formal education settings. Its core premise is that when people claim to hold certain values but appear to be acting in ways that directly contravene those values, the major problem is a lack of clear consensus or shared understanding about what the values might 'look like' when translated into action, within a given practical context. This lack of clarity can result in the values being over-ruled by factors such as convenience, money and time. This is exactly where intersubjective 'indicator' design processes can be helpful – not necessarily for evaluating anything or assessing anyone, but for encouraging different stakeholders to share their respective insights and to generate a common understanding of what matters to them as a group in the form of a set of statements or principles. These statements can support individuals and groups in standing up for their values in the face of competing pressures. In the paper, I used the term 'skill statements' (rather than indicators) in order to focus students' and teachers' attention on the concept of sustainability-related competencies, which I will explore more fully in the discussion of Paper 4 below.

6.5.2. Limitations and critiques

In common with Paper 1, this paper has a pragmatic rather than a theoretical starting point and is inductive, not deductive, in nature. While acknowledging relevant literature as a backdrop to the research, it does not use this literature as a source of theoretical frameworks or code books that can be used systematically to solve problems. In section 5.1, for example, I have stated that the paper will not `attempt the challenge of identifying, and proposing strategies for removing, each separate barrier to pro-environmental behaviour', but, instead, will 'present a holistic solution that has already proven helpful for identifying and bridging value-action gaps' in CSOs and 'demonstrate that this solution can be adapted for use in schools'.

What has become evident with hindsight is that the tone of Chapter 5 is, in certain places, more redolent of a marketing professional seeking to sell their particular 'solution' to prospective clients than of an academic exploring an intriguing research question. While this tone is not entirely out of character with the nature of the book in which this chapter was published, nor with the subject matter being researched (which is inherently inductive), it may not inspire confidence in the critical reader who might reasonably demand to know how the research fits into broader landscapes of knowledge creation.

In Chapter 5 I attempted to circumvent any criticism on the grounds of insufficient consideration of the relationships between values and behaviour, or inadequate reference to theories of behaviour change in general, by bypassing the majority of this literature and instead framing the work as an example of 'Research through Design' (RtD). I explain on page 120 of the published book, for example, that RtD seeks to transform a problematic situation to a preferred situation through the creation of 'a redesigned thing' that simultaneously functions as a contribution to knowledge and a contribution to society. I then proceed to state three pairs of propositions underlying the research, each one contrasting a current situation which is 'problematic' (presumably from my own viewpoint, as I neither unpack the definition of 'problem', nor provide any academic or practitioner-led justification for describing the respective situations in this way) with a prospective future situation that would be 'preferred' (again, presumably by me). Yet the RtD section is itself poorly theorised, both from the standpoint of RtD-specific literature and in the broader sense of how the

epistemology of design research is distinct from the respective epistemologies of research in the sciences and arts (Cross, 1999). I have rectified the former in the introduction to this thesis, and will now focus on the latter.

It is in distinguishing design research from other forms of research that the theoretical lens of Nicolescuian transdisciplinarity, discussed in relation to Chapter 3 above, can again provide useful insights. We can see that the table of comparisons between the epistemology of design research and that of the sciences and arts, which I have adapted from Cross (1999) and included as Table 5-1 on page 176, maps neatly on to the diagram used by McGregor and Volckmann (2013) to explain the intersection of the 'external' and 'internal' levels of Reality in the form of the 'hidden third' or 'zone of non-resistance'. Where Cross talks about the sciences and arts intersecting in the arena of design, McGregor and Volckmann illustrate the 'external' levels of Reality (described respectively as 'environmental, economic, and planetary/cosmic', and the 'internal' levels ('individual, social/political, philosophical, historical') intersecting in the arena of 'religions, spiritualities, cultures/art'. It is precisely this grey area or interface between the internal and external levels of Reality that characterises the 'missing pillar' of sustainability, which constitutes the focus of Chapter 3.

This leads to the insight that what Cross (2011) has termed 'designerly ways of knowing' can also be described as *transdisciplinary* ways of knowing, and that all transdisciplinary researchers and practitioners are, in some sense, 'co-designers' of their own preferred-future scenarios[21]. Beyond this, design itself can be recognised as a separate means of creating 'zones of non-resistance', distinct from culture, art, religions and spiritualities. It is through design, and by design, that we can initiate conversations that interweave the epistemologies of the sciences and humanities.

---

[21] This generalisation does not detract in any way from the importance of promoting transdisciplinary *design* in a purposeful and conscious way, e.g. by exploring the ways in which the axioms of transdisciplinarity can be applied directly to design processes. This is exemplified by the pan-European 'Transdisciplinary Design Network', founded by Tuba Kocaturk and Richard Koeck in partnership with the Design Council: see https://www.liverpool.ac.uk/architecture/news/reports/transdisciplinary/ (accessed 14 December 2018).

More specific critiques of Paper 3, all relating to the research and the paper itself rather than the approach, are as follows:

(a) The central research study, as published, is not replicable. In spite of describing it as 'a new analysis of several datasets', which were collected during 'explorations of values in educational contexts' (page 121), I provided very little detail about how the original datasets were obtained, where they might be found, or what type of data they contained. While this problem is shared to some extent by Paper 1, it is less of an issue in that case because the original research question (namely, 'what were the outcomes of the ESDinds project field visits?') was well-defined, with a clarification that the interviewees were not primed or prompted to mention any specific benefits. In Paper 3, I neither stated the actual research questions that were used to generate the three different datasets from Ireland, the UK and Tanzania, nor mentioned the key point that there were deliberate commonalities between them. All of them were, in fact, variations on the question of 'what teachers find most meaningful, valuable or worthwhile about their work in education').

(b) In a similar vein, when talking about the written survey that was subsequently administered to PERL workgroup members (p. 125), I stated only that it was 'modelled on the "Educate Together" questionnaire described above, but with a stronger EfSRL focus', without realising that I had neither described the questionnaire nor explained the meaning of 'Educate Together'. This was missed by the reviewers, who accepted the paper without revisions. For clarification, 'Educate Together' was the name of the NGO which organised the 2011 conference in Dublin, Ireland where I gathered data on participants' values, through the administration of a survey questionnaire in conference packs. The survey had several questions, including asking participants to recall an experience of education that had been particularly positive or memorable for them, as well as asking them about what they had found particularly meaningful, valuable or worthwhile in relation to their work in education. Including the survey as an appendix to the paper would have improved the recoverability and hence the quality of the Research through Design work, as well as providing a useful tool for

future researchers who might wish to replicate the work in other cultural contexts or extend it in new directions.

(c) The selection of data collection sites was convenience-based, rather than chosen for maximum diversity sampling, and thus reflects a strange geographical bias – two locations in Anglophone western Europe (UK and Ireland) and one in East Africa (Tanzania). This arose from the fact that the project had severe budget limitations, and the research design thus represented a compromise between rigour and practicability. It would ideally have been more appropriate to disseminate questionnaires and/or identify interviewees through a global network such as the Earth Charter Initiative, in order to ensure the maximum geographical spread of responses – with particular attention paid to Indigenous populations and minority groups. The lack of cultural and geographical diversity in the original datasets renders the findings of the meta-analysis largely meaningless, in terms of being able to draw valid conclusions about what matters most to educators around the world. They were, however, useful for creating 'trigger statements' that served as prompts for reflection within the contexts where they were tested – one school in the UK, described in the chapter, and another in Finland.

(d) The paper does not include any examples of the types of 'skill statements' that were included in the original draft toolkit, which were derived from the Research through Design process, but only those that were added by the students themselves during the capacity-building sessions (p. 127). There are not even any details of what the `draft toolkits' actually are, or how to access them, which severely limits the usefulness of the paper for practitioners who might have been interested in exploring this area of work. For reference, the three draft toolkits produced as a result of the PERL project were:

(i) *Measuring What Matters: Values-Based Indicators, A Methods Sourcebook,* targeted primarily to the leaders and governors of secondary schools (available at https://iefworld.org/fl/PERL_toolkit1.pdf);

(ii) *Discovering What Matters: A Journey of Thinking and Feeling: Activities Designed With Students, For Students*, which is the 'draft toolkit' referred to in Paper 3 and includes activities for secondary school students and teachers, https://iefworld.org/fl/PERL_toolkit2.pdf;

(iii) *Growing a Shared Vision, A Toolkit for Schools*: *Activities for Organisational and Staff Development*, found at https://iefworld.org/fl/PERL_toolkit3.pdf.

A broader criticism of the work presented in Chapter 5 is that at the time of writing, in 2015, I had spent very little time as an educator in school contexts – no more than the occasional arts workshop or guest speech in assemblies – and had only a very limited awareness of UK National Curriculum requirements or the other constraints faced by teachers. I was conscious of this limitation at the time, and attempted to address it through ongoing dialogue with lecturers in the School of Education at the University of Brighton. Their initial feedback was that the draft toolkit was too long, too academic in its tone, and generally too overwhelming to be useful for busy teachers who would not automatically know what to do with it. This resulted in splitting it into the three separate documents listed above: *Discovering What Matters,* which may have some value (in either classroom or School Council contexts, or both) but still requires further testing and refining, and *Measuring What Matters* and *Growing a Shared Vision,* which have not been tested in any context and are unlikely to be fit for purpose. Indeed, the respective purposes of these two latter documents were never clearly stated, although the former is presented as a 'methods sourcebook' and the latter as a 'toolkit for staff and governors'. Having recently taken up a part-time post as a science teacher and policy adviser in a small independent school in southern England, I can now more fully understand the criticisms levelled against the initial versions of the draft toolkits. On the basis of my own experience of teaching, I am now in a better position to develop resources that meet the needs of teachers and school governors in relation to supporting transformational learning for sustainability.

Furthermore, I recognise that the independent school sector in the UK, in particular, is in a strong position to develop sustainability skills and leadership and even, potentially,

transdisciplinary research and design skills. This is on account of recent advances in the inspection and assessment of independent schools, announced during a recent seminar organised by the UK Government's Office for Standards in Education (OFSTED), which allow inspectors to judge each 'non-association' independent school according to its own criteria. Provided that national minimum standards are met, the inspector is expected to assign a grading not by adhering to a particular ideology or theory of education, but by examining the school's own vision, values, self-evaluation and school improvement plan. Schools can choose whether to assess students in relation to national curriculum targets, their own framework of aims and objectives, or both (OFSTED, 2018, personal communication at a training seminar held in London).   This allows them a considerable degree of freedom to experiment and innovate.

6.6. Cross-cutting reflections applicable to all four published works

It was recognised during the ESDinds project that there was something crucially important about presenting participants with a set of prompts for discussion, after the initial values elicitation phase. These were variously framed as 'indicators', 'proto-indicators', 'menu items', 'values-based statements' or 'trigger statements'. The prompts were presented in a variety of formats during the course of the ESDinds project and the subsequent 'Starting from Values' project – a printed checklist, a booklet with the statements shown in speech bubbles, or a set of individual cards, stickers or paper strips that the participants were asked to sort into piles. However, the reasons for the central importance of this activity were not fully understood at the time when the earlier papers were written. Accordingly, both the significance of what I have recently termed the 'challenge phase' and the mechanism of its operation were under-emphasised in the original publications.

I referred to the challenge phase initially as a process of 'reflection on an indicator list' (Chapters 2-3), and subsequently as 'gently disruptive probing to force conscious recognition of tacit knowledge [and] to denaturalise socially indicated responses' (Chapter 4).   In all the papers, however, I missed the key point that the *purpose* of this stage of the process was to challenge participants not only to leave their respective mental 'comfort zones' and become

aware of social desirability bias, but also to empathise with other perspectives - the very foundation of both intersubjectivity and transdisciplinarity.

What has since become clear is that there may be other means of achieving the same end, besides the provision of printed artefacts. These could include, for example, the presentation of short films, artworks or multimedia materials that introduce different viewpoints or insights, or interactions with a wider community responding directly to the elicited statements (either in real time, or over a period of hours or days) through an online platform.

Subsequent research and reflection has led me to make the following recommendations in relation to materials used in the challenge phase, regardless of their format:

- They should be sufficiently provocative to push participants to re-examine their comfortable assumptions, not only about what matters to them, but about what can safely be articulated in the group space without fear of repercussions. This may be especially important in relation to 'triple-T' learning, where participants may initially be nervous about voicing opinions that transgress social norms, cross disciplinary boundaries, or relate to a different 'level of Reality' from the one that is customarily referenced within the group.
- At the same time, the materials should use vocabulary that feels familiar and natural to participants. If the language used in the materials feels too far removed from the day-to-day reality of the working group, or if it causes offence or triggers negative emotional responses, members may disengage from the process or actively disrupt it.
- The materials should facilitate the realisation that the diverse participants of the group may have different, or even competing, values and priorities.

6.7. Towards a research agenda for the future

Reflecting on the four publications as a body of work highlights the fact that on one level, the creation of inductive and intersubjective 'indicators' (or skill statements) can variously be understood as a transdisciplinary design process, a transdisciplinary research

process, and/or a process of Research through Design with theoretical roots in Nicolescuian transdisciplinarity. On another level, it constitutes a response to the challenge of *how to evaluate the societal impacts* of local or glocal transdisciplinary projects relating to sustainability – especially those impacts that are 'important but not easily measurable' (Lang, 2012). Such projects typically play out within overlapping contexts of research, industry and education, transcending the boundaries between these domains just as they transcend the outdated idea of discrete and clearly bounded academic disciplines.

The productive tension between 'understanding the process' and 'applying the process' of inductive / intersubjective indicator creation can be seen in all of the papers discussed in this thesis, as well as in my more recent work focusing on evaluating multi-stakeholder collaborative research in the arts and humanities (Brigstocke et al., 2017; Brockwell, 2018; Harder & Burford, 2019). In each of these situations, the indicator design process itself is transdisciplinary in nature; but it is also nested within a larger context of transdisciplinary research and praxis that draws attention to the intimate relationship between education and evaluation, and the necessity of questioning the values, priorities and mindsets that underlie the practice of evaluation design. In the final section of this thesis, I will use these points as the foundations of a future research agenda.

6.7.1. Evaluation design and meta-evaluation as an emerging research arena

It is widely recognised that *what matters is what's measured*, in the sense that teachers are often accused of 'teaching to the test', and students engaged in revising for examinations will refuse to spend their time learning material that is unlikely to be examined (c.f. Henshaw). Similarly, in community development contexts, the tendency is always to prioritise those outputs and outcomes that are measured in the final project evaluation - whether or not these actually represent the desired long-term impact. As I have previously pointed out in Chapter 4, and in the introduction to this thesis, the 2015 Millennium Development Goals Report explicitly highlights the fact that 'what gets measured gets done' (United Nations, 2015, p. 10). In this sense, the individuals or organisations responsible for setting examinations, identifying SDG indicators, and deciding the criteria against which projects are assessed – not the evaluators themselves, but the evaluation *designers* - have a

vital and often overlooked role in determining the future direction(s) of society. Their authority is largely unquestioned, in that educators, examiners and policy-makers alike are all too often failing to ask the deeper question of *whether what's measured is what really matters.*

This, of course, begs the question 'matters to whom?' – if we acknowledge that different stakeholders have differing values and priorities, and that automatically deferring to the most powerful and vocal stakeholder is not the only viable strategy. However, there may be some questions that are so vital for the future of humanity and the Earth, especially in these challenging times, that they transcend individual and group value differences. Re-orienting evaluation design around such questions, to the extent that this can be done in isolation from party politics and related power games, could be a subtle but effective way of creating positive change. Evaluation designers could enquire, for example, whether the examination grades received by students (or the outputs created by project participants) indicate mastery of all the specific sustainability competencies described in the model in Chapter 4. Alternatively, it would be possible to investigate whether students' grades or participants' project outputs reflect their respective abilities to understand and interact with others in ways that are rooted in love rather than fear; to collaborate and co-design solutions to mutual problems; to be resilient and flexible in times of rapid change; to connect in meaningful ways with the natural world; to respect what Indigenous communities hold sacred; and to build communities that are ecologically, economically and socially sustainable.

I am aware, of course, that in attempting to frame such questions – even in the most speculative way - I am yet again seeing through the filter of my own biases and blind spots. These may be very different now from those that I held in 2011, when I first began the programme of research that led ultimately to these four publications; but they are still no less problematic. The questions that I have set out in the preceding paragraph are deeply rooted in my personal views on 'what matters' from a global perspective, and have been included only as examples. The themes and questions set out below could form the core of a research agenda on collaborative evaluation design and what could be termed 'meta-evaluation' (i.e. 'the evaluation of evaluation') that seeks to transpose this discussion from a subjective level to an intersubjective one.

Theme 1: Contexts

A key theme within a research agenda for collaborative evaluation design is how to engage people in 'genuine dialogue' of the type described by Buber (1979), namely a dialogue that is grounded in the recognition of difference and the quest for authentic relationship (see also Talamo & Pozzi, 2011).

As noted in the Introduction, the establishment of genuine dialogue is fundamental to building consensus on what is valued by all stakeholders in the process of designing evaluations or assessments. At its heart is the willingness to interrogate the stories that we are accustomed to telling ourselves about what matters and why we are here, and in particular, to challenge those aspects of our received narratives that have closed our ears to other perspectives. Our existing stories may exclude the sacred and spiritual elements that are key dimensions of Indigenous world-views; they may be neo-colonialist, racist, misogynist, heteronormative or cisnormative; and they may be deeply anthropocentric, either in the sense of neglecting other species altogether, or assuming that only species that are perceived by humans as useful and/or beautiful are worthy of our time and attention. These flaws in our stories are not obvious to us from the start: they become evident only when we, as individuals with one or more dimensions of privilege, interact with members of the respective marginalised groups and realise that their own narratives are very different.

Some relevant research questions are as follows:

1.1. What are the contexts in which people are already engaging in 'genuine dialogue' in relation to evaluation design – i.e. creating indicators, assessable learning outcomes, marking criteria, assessment tools, etc. that draw on the values and perspectives of multiple stakeholders, rather than being imposed from the top down?

1.2. What are the similarities and differences between the evaluation design processes that are currently being employed in different contexts – e.g. examining boards,

universities, non-formal education settings, glocal networks, and multi-stakeholder research projects? To what extent are any of them inductive and intersubjective?

1.3.   How might stakeholders from these different contexts share methods and experiences with one another, using 'cross-pollination' to create new hybrid approaches to evaluation design that combine their respective advantages?

Theme 2: Facilitation and Stimulus Materials

The process of inductive / intersubjective evaluation design demands careful facilitation, whether through a face-to-face approach such as the *WeValue* workshops described in Paper 1, or an interaction mediated by online collaboration tools. Some possible research questions relating to this theme are as follows:

2.1.   How can facilitators acknowledge their own privilege and create a safe space in which members of marginalised groups, who may have been traumatised by previous experiences of abuse and discrimination, feel comfortable enough to share their stories and insights openly and without fear? How can the inhibitory effects of social desirability bias, which could result in a tendency to edit or water down narratives to make them more palatable for 'mainstream' audiences, be overcome?   Does a guarantee of anonymity help or hinder full disclosure?   How useful is indirect questioning?

2.2.   To what extent is it helpful to seek maximum diversity in the composition of groups, and how can facilitators ensure that all relevant stakeholders are invited to the table?

2.3.   Which ESD competencies and/or personal qualities do facilitators need in order to be effective in facilitating 'genuine dialogue' without dominating the conversation?   Which do they need in order to be successful in steering discussions in a useful direction and enabling participants to co-create the desired deliverable(s), such as a list of indicators, a set of criteria or an evaluation plan, on time and on budget?

2.4.   To what extent can the arts, or artists as 'entrepreneurs in conventions', contribute to the critical examination of received narratives and the creation of zones of non-resistance between groups or individuals with different perspectives?

Theme 3: Outcomes

Following the preliminary study outlined in Chapter 2, it is vitally important to document not only the processes, but also the outcomes of intersubjective / inductive evaluation design in practice. This is not only an example of Research through Design, but also of 'meta-evaluation' – evaluating the outcomes and longer-term impacts of evaluations themselves, and relating them to evaluation design. The research programme that has recently been implemented by the UK's Office for Standards in Education (Ofsted) provides a useful example of meta-evaluation. Preliminary findings showing that Ofsted inspections have not only increased teacher workload to a point at which it is affecting wellbeing, but also distracted headteachers from the curriculum and in some cases led them to employ consultants to conduct 'mocksted' inspections, has led to a proposal for revision of the criteria against which UK primary and secondary schools are assessed (Ofsted, 2018).

Some questions that could form the basis of a research agenda on outcomes include:

3.1. How does participating in a process of inductive / intersubjective evaluation design change the professional relationships between participants? To what extent, if at all, does participants' increased understanding of different values and world-views impact on their collaborative initiatives with different partners (e.g. a large funder employing a new approach to project evaluation with its other grantees, or a policy-maker changing their strategic direction)?

3.2. Does the introduction of new indicators, criteria, assessable learning outcomes and/or assessment tools lead to sustained positive change in the collectively-agreed direction(s), as evidenced by a continuous improvement in grades or scores? Do educators, students or project participants spontaneously change their practice in order to meet the new criteria or score highly on the new assessments, without being directly instructed to do so?

3.3. What are the wider impacts of inductive or intersubjective evaluation design in society? Can redesigning evaluations and assessments effect a change in worldview, or a shift in what is valued within an organisation, local community or glocal network, as implied by the term 'conceptual use of indicators'?

6.8. Concluding Remarks

In this thesis, I have demonstrated the potential of inductive and intersubjective approaches to indicator design and assessment tool design in Education for Sustainability. Notwithstanding the potential for misuse and abuse, which can be mitigated through careful attention to the composition of the participant group and through skilful facilitation, I have illustrated that this area of research has the potential to bring about significant and lasting transformation in the ways in which educational initiatives are designed, implemented, and – crucially – monitored and evaluated. It is particularly well suited to the types of educational initiatives that are the most challenging to evaluate using conventional methods – namely, those that are transformative, transgressive and transdisciplinary ('triple-T learning').

The current political climate, at the time of writing, is characterised by societal norms in the UK, the USA and several European countries that appear to be moving in the worrying direction of combining materialistic consumerism, neo-fascism, anti-scientism, and climate change denial. For this reason, the effective design of educational initiatives that empower students to transgress societal norms and transform oppressive structures may be humanity's best hope, or perhaps our *only* hope, of maintaining the Earth's climate within habitable boundaries through the twenty-first century and beyond.   Viewed within this context, establishing initiatives to create curricula and teacher training programmes that enhance

transformative and transgressive learning, in both formal and non-formal settings, should be among the highest priorities of all who are genuinely concerned with the promotion of education for sustainability.

To prevent the artificial boundaries between academic disciplines, professional specialities and wider civil society from impeding the progress of these vital initiatives, the third 'T' – transdisciplinarity – is no less important. The shift away from 'either/or' and towards 'both/and' logic – the logic of the included middle described by Nicolescu (1998, 2002, 2010) – is one of the key aspects of this transition. It places not only design, but also the arts, culture and spirituality - the various components of the 'missing pillar' of sustainability that I identify in Chapter 3, which may ultimately be the foundation underpinning all of the other 'pillars' – firmly at the centre of education, rather than viewing them as optional extras.

The research agenda that I have set out in this concluding chapter is an ambitious one, which would require collaboration between a diverse range of stakeholders. It is likely to be of particular interest to Indigenous communities and their allies, as well as to the transdisciplinary, collaborative and transformative design and research networks that are already making advances in 'triple-T' learning. In addition, as I highlighted in a recent essay – the first that I have published in my new name (Brockwell, 2018) – it is likely to be welcomed by artists and academic researchers who are either already engaging, or would like to engage, in collaborative research. It is my hope that this thesis will serve as a springboard for similar work, and that new contexts will continue to be discovered as these research agendas evolve.

Ashley Jay Brockwell (formerly Gemma Burford)

December 2018

References

Abram, D. (1996). *The spell of the sensuous: perception and language in a more-than-human world*. New York: Pantheon.

Agle, B. R., & Caldwell, C. B. (1999). Understanding research on values in business: A level of analysis framework. *Business & Society, 38*(3), 326-387.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Allport, G. W., Vernon, P. E., & Lindzey, G. (1951). *A study of values*. Boston, MA: Houghton Mifflin.

Alvesson, M., & Sveningsson, S. (2003). Good visions, bad micro-management and ugly ambiguity: contradictions of (non-)leadership in a knowledge-intensive organization. *Organization studies, 24*(6), 961-988.

AMA. (2002). *American Management Association 2002 Corporate Values Survey*. Retrieved from New York: http://www.amanet.org/training/whitepapers/2002-Corporate-Values-Survey-35.aspx

Andrews, M. C., Baker, T., & Hunt, T. G. (2011). Values and person-organization fit: Does moral intensity strengthen outcomes? *Leadership & Organization Development Journal, 32*(1), 5-19.

Archer, B. (1995). The nature of research. *CoDesign, January 1995*, 6-13.

Argyris, C., & Schon, D. A. (1978). *Organizational learning*. Reading, MA: Addison-Wesley.

Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal, 24*(2), 377-385.

Artto, K., Kulvik, I., Poskela, J., & Turkulainen, V. (2011). The integrative role of the project management office in the front end of innovation. *International Journal of Project Management, 29*(4), 408-421. doi:10.1016/j.ijproman.2011.01.008

Attaran, A. (2005). An immeasurable crisis? A criticism of the Millennium Development Goals and why they cannot be measured. *PLoS Medicine, 2*(10), 955-961.

Babbie, E. R. (2010). *The practice of social research* (International edition of 12th edition ed.). Belmont, CA: Thomson / Wadsworth.

Baha'i International Community. (1998). *Valuing spirituality in development: Initial considerations regarding the creation of spiritually based indicators for development. A concept paper written by the Baha'i International Community for the World Faiths and Development Dialogue, Lambeth Palace, London, 18-19 February 1998.* London: Baha'i Publishing Trust.

Baha'i International Community. (2002). *Sustainable development: the spiritual dimension. A statement to the First Session of the Preparatory Committee of the World Summit on Sustainable Development*. Retrieved from New York: http://bic.org/statements-and-reports/statements/sustainable-development-the-spiritual-dimension

Bahm, A. J. (1994). *Ethics: the science of oughtness*. Amsterdam, The Netherlands: Rodopi B.V.

Bardi, A., Lee, J. A., Hofmann-Towfigh, N., & Soutar, G. (2009). The Structure of Intraindividual Value Change. *Journal of Personality and Social Psychology, 97*(5), 913-929. doi:10.1037/a0016617

Barrass, S. (2016). An annotated portfolio of research through design in acoustic sonification. *Leonardo, 49*(1), 72-73.

Bell, S., & Morse, S. (2008). *Sustainability Indicators: Measuring the Immeasurable?* London, UK: Earthscan.

Bell, S., & Morse, S. (2011). Sustainable development indicators: the tyranny of methodology revisited. *Consilience: the journal of sustainable development, 6*(1), 222-239.

Berg, P. G., & Nycander, G. (1997). Sustainable neighbourhoods—a qualitative model for resource management in communities. *Landscape and Urban Planning, 39*(2–3), 117-135. doi:http://dx.doi.org/10.1016/S0169-2046(97)00050-9

Biermann, F. (2007). 'Earth system governance' as a crosscutting theme of global change research. *Global Environmental Change, 17*, 326-337.

Biggs, J. B. (1985). The role of meta-learning in study process. *British Journal of Educational Psychology, 55*, 185-212.

Biggs, M. A., & Buchler, D. (2007). Rigor and practice-based research. *Design Issues, 23*(3), 62-69.

Blake, J. (1999). Overcoming the 'value-action gap' in environmental policy: Tensions between national policy and local experience. . *Local Environment, 4*(3), 257-278.

Blizzard, J. L., & Klotz, L. E. (2012). A framework for sustainable whole systems design. *Design Studies, 33*, 456-479.

Bourdieu, P. (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press.

Bradbury Huang, H. (2009). What is good action research? Why the resurgent interest? *Action Research, 8*(1), 93-109.

Braithwaite, V. A., & Law, H. G. (1985). Structure of Human Values: Testing the Adequacy of the Rokeach Value Survey. *Journal of Personality and Social Psychology, 49*(1), 250-263. doi:doi:10.1037/0022-3514.49.1.250

Brigstocke, J., Hoover, E., Harder, M. K., Graham, P., De Sousa, S., Dearden, A., . . . Foskett, C. (2017). Implicit values: uncounted legacies. In K. Facer & K. Pahl (Eds.), *Valuing interdisciplinary collaborative research: beyond impact* (pp. 65-83). Bristol, UK: Policy Press.

Brockwell, A. J. (2018). Values-focused evaluation: measuring what matters most. In S. Leighton & P. Mitchell (Eds.), *The Exchange: revealing collaborative values*. London: The Culture Capital Exchange.

Brown, B., Buchanan, R., DiSalvo, C., & Margolin, V. (2013). Introduction. *Design Issues, 29*(2), 1-3.

Brown, B., Buchanan, R., DiSalvo, C., & Margolin, V. (2014). Introduction. *Design Issues, 30*(2), 1-3.

Brown, D., & Crace, R. K. (1996). Values in Life Role Choices and Outcomes: A Conceptual Model. *The Career Development Quarterly, 44*, 211-223.

Buber, M. (1979). *Ich und Du. Das Dialogische Prinzip [I and Thou: The Dialogic Principle]*. Heidelberg, Germany: Verlag Lambert Schneider.

Buchanan, R. (1995). Wicked problems in design thinking. In V. Margolin & R. Buchanan (Eds.), *The Idea of Design* (pp. 14). Cambridge, MA: The MIT Press.

Buchko, A. A. (2007). The effect of leadership on values-based management. *Leadership & Organization Development Journal 20*(1), 36-50.

Burford, G. (2015). Collaborative research for sustainability: an inside-out design manifesto. *CASIC Working Paper Series: https://www.keele.ac.uk/casic/workingpaperseries/,* (2).

Burford, G., Hoover, E., Dahl, A., & Harder, M. K. (2015). Making the invisible visible: designing values-based indicators and tools for identifying and closing 'value-action gaps'. In V. W. Thoresen, R. J. Didham, J. Klein, & D. Doyle (Eds.), *Responsible living: concepts, education and future perspectives* (pp. 113-134). Cham, Switzerland: Springer International.

Burford, G., Hoover, E., Jarvis, D., & Harder, M. K. (2014). Assessing group learning processes: reflections on using values-based indicators as assessment criteria with final-year undergraduates In G. Wisker, L. Marshall, S. Greener, & J. Canning (Eds.), *Evolving Experiences: Articles from the Learning and Teaching Conference 2013* (pp. 58-67). Falmer, UK: University of Brighton.

Burford, G., Hoover, E., Stapleton, L., & Harder, M. K. (2016). An unexpected means of embedding ethics in organizations: preliminary findings from values-based evaluations. *Sustainability, 8*(7), 612.

Burford, G., Hoover, E., Velasco, I., Janouskova, S., Jimenez, A., Piggot, G., . . . Harder, M. K. (2013). Bringing the 'missing pillar' into Sustainable Development Goals: towards intersubjective values-based indicators. *Sustainability, 5*, 3035-3059.

Burford, G., Kissmann, S., Rosado-May, F. J., Alvarado Dzul, S. H., & Harder, M. K. (2012). Indigenous participation in intercultural education: learning from Mexico and Tanzania. *Ecology and Society, 17*(4), 33.

Burford, G., Tamas, P., & Harder, M. K. (2016). Can we improve indicator design for complex Sustainable Development Goals? A comparison of a values-based and conventional approach. *Sustainability, 8*(9), 861.

Burford, G., Velasco, I., Janouskova, S., Zahradnik, M., Hak, T., Podger, D., . . . Harder, M. K. (2013). Field trials of a novel toolkit for evaluating 'intangible' values-related dimensions of projects. *Evaluation and Program Planning, 36*(1), 1-14. doi:http://dx.doi.org/10.1016/j.evalprogplan.2012.04.005

Burger, P., & Kamber, R. (2003). Cognitive integration in transdisciplinary science: knowledge as a key notion. *Integrative Studies, 21*, 43-73.

Cameron, W. B. (1963). *Informal sociology: a casual introduction to sociological thinking*. New York: Random House.

Carroll, A. B. (2000). Ethical challenges for business in the new millennium: corporate social responsibility and models of management morality. *Business Ethics Quarterly, 10*(1), 33-42.

Carteron, J.-C., & Decamps, A. (2014). *The sustainability literacy test: can universities be sure they are producing sustainability literate graduates? One year report presented on the occasion of the World Conference on Education for Sustainable Development*. Retrieved from Nagoya, Japan: https://sustainabledevelopment.un.org/content/documents/1592Literacytestreport.pdf

Cebrián, G., & Junyent, M. (2015). Competencies in Education for Sustainable Development: Exploring the Student Teachers' Views. *Sustainability, 7*(3), 2768.

Cha, S. E., & Edmondson, A. C. (2006). When values backfire: Leadership, attribution, and disenchantment in a values-driven organization. *Leadership Quarterly, 17*(1), 57-78.

Christensen, T., & Lægreid, P. (2011). Ethics and Administrative Reforms. *Public Management Review, 13*(3), 459-477. doi:10.1080/14719037.2011.553320

Christofi, A., Christofi, P., & Sisaye, S. (2012). Corporate sustainability: historical development and reporting practices. *Management Research Review, 35*(2), 157-172.

Clegg, S., Kornberger, M., & Rhodes, C. (2007). Business Ethics as Practice. *British Journal of Management, 18*(2), 107-122. doi:10.1111/j.1467-8551.2006.00493.x

Clugston, R. (2011). Ethical Framework for a Sustainable World: Earth Charter Plus 10 conference and follow up. *Journal of Education for Sustainable Development, 5*(2), 173-176.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education* (7th ed.). Abingdon, UK: Routledge.

Combes, B. P. (2005). The United Nations decade of education for sustainable development (2005–2014): learning to live together sustainably. *Applied Environmental Education and Communication, 4*(3), 215-219.

Conrad, J. (2002). Limits to interdisciplinarity in problem oriented social science research. *The Journal of Transdisciplinary Environmental Studies, 1*(1), 1-15.

Convertino, C., Brown, A., & Wilson, M. A. F. (2015). Public borders, private crossings: Anthropological approaches to neoliberalism in US educational policy and spaces. *Policy Futures in Education, 15*(2), 135-140. doi:10.1177/1478210315622914

Corcoran, P. B., Vilela, M., & Roerink, A. (2005). *The Earth Charter in Action: Toward a Sustainable Development.* Amsterdam, The Netherlands: Royal Tropical Institute (KIT).

Cousins, J. B. (2007). Process use in theory, research and practice. *New Directions for Evaluation, 116*.

Crenshaw, K. (1989). Demarginalising the intersection of race and sex: a Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 139-168.

Cross, N. (1995). Editorial. *Design Studies, 16*, 2-3.

Cross, N. (1999). Design research: a disciplined conversation. *Design Issues, 15*(2), 5-10.

Crossley, N. (1996). *Intersubjectivity: the fabric of social becoming*. London: Sage Publications.

Dahl, A. (2012). Achievements and gaps in indicators for sustainability. *Ecological Indicators, 17*, 14-19. doi:doi:10.1016/j.ecolind.2011.04.032

Dahl, A., Harder, M. K., Mehlmann, M., Niinimaki, K., Thoresen, V., Vinkhuyzen, O., . . . Velasco, I. (2014a). *Discovering what matters: a journey of thinking and feeling. Activities developed with students, for students. PERL Values-Based Learning Toolkit 2.*: Partnership for Education and Research about Responsible Living (PERL).

Dahl, A., Harder, M. K., Mehlmann, M., Niinimaki, K., Thoresen, V., Vinkhuyzen, O., . . . Velasco, I. (2014b). *Growing a shared vision: a toolkit for schools. Activities for organizational and staff development. PERL Values-Based Learning Toolkit 3.*: Partnership for Education and Research about Responsible Living (PERL).

Dahl, A., Harder, M. K., Mehlmann, M., Niinimaki, K., Thoresen, V., Vinkhuyzen, O., . . . Velasco, I. (2014c). *Measuring what matters: values-based indicators. A methods sourcebook. PERL Values-Based Learning Toolkit 1*: Partnership for Education and Research about Responsible Living (PERL).

Dalal-Clayton, B., & Bass, S. (2002). *Sustainable development strategies: a resource book* London: Earthscan.

Davidov, E. (2010). Testing for comparability of human values across countries and time with the third round of the European Social Survey. *International Journal of Comparative Sociology, 51*(3), 171-191. doi:10.1177/0020715210363534

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in: Testing the adequacy of the European Social Survey to measure values in 20 countries. . *Public Opinion Quarterly 72: , 420-445.*

De Leo, J. (2012). *Quality education for sustainable development: an educator handbook for integrating values, knowledge, skills and quality features of Education for Sustainable Development in schooling*. Adelaide, Australia: UNESCO APNIEVE Australia Publishing.

Debats, D. L. H. M. (1996). The structure of human values: a principal components analysis of the Rokeach Value Survey (RVS). In *Meaning in life : psychometric, clinical and phenomenological aspects* (pp. 47-55). Groningen: De Rijksuniversiteit Groningen

Delors, J., Al Mufti, I. a., Amagi, I., Carneiro, R., Chung, F., Geremek, B., . . . Zhou, N. (1996). *Learning: the treasure within. Report to UNESCO of the International Commission on Education for the Twenty-First Century*. Retrieved from Paris, France:

Dixon, B. R. (1978). An investigation into the use of Raths' values clarification strategies with grade eight pupils (Doctoral dissertation, Michigan State University, 1978). *Dissertation Abstracts International, 39*, 5987A.

Dobson, A. (1996). Environmental sustainabilities: an analysis and a typology. *Environmental Politics, 5*(3), 401-428.

Driskell, D. (2002). *Creating Better Cities with Children and Youth: A Manual for Participation*: ERIC.

Druin, A. (2005). What Children Can Teach Us: Developing Digital Libraries for Children with Children1. *The Library, 75*(1).

Du, S., Bhattacharya, C. B., & Sen, S. (2010). Maximising business returns to corporate social responsibility (CSR): the role of CSR communication. *International Journal of Management Reviews*.

Ebrahim, A. (2002). Information struggles: The role of information in the reproduction of NGO-funder relationships. *Nonprofit and Voluntary Sector Quarterly, 31*(1), 84-114.

Ebrahim, A. (2003). Accountability in practice: Mechanisms for NGOs. *World Development, 31*(5), 813-829.

ECI Secretariat. (2010). *Earth Charter Initiative Handbook*. Costa Rica: Earth Charter Initiative.

ECI Secretariat. (2011). *Evaluation of e-GLO3 Earth Charter Global Learning Opportunity*. Retrieved from San José, Costa Rica:

Education for All. (2014). *Teaching and learning: achieving quality for all. Education for All 2013-14 global monitoring report*. Retrieved from Paris:

Elliott, J. A. (2012). *An introduction to sustainable development*. London: Routledge.

Erpenbeck, J., & von Rosenstiel, L. (Eds.). (2003). *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis [Handbook of competency measurement. Recognising, understanding and valuing competencies in professional, pedagogical and psychological practice]*. Stuttgart, Germany: Schäffer, Poeschel.

ESDinds. (2010a). *Deliverable 11: Findings from Set 1 Sustainable Development Indicators Research*. Retrieved from Brighton, UK:

ESDinds. (2010b). *ESDinds working paper: Report of field visit to organisation 'HV' for testing of Set 1 Sustainable Development Indicators*. Brighton, UK: University of Brighton.

ESDinds. (2010c). Face-to-face interview with 'Carlos', lead ESDinds researcher for field visit to organisation 'DB'. In G. Burford (Ed.), *ESDinds working paper: Transcripts of anonymised semi-structured interviews with participants in field tests of Set 1 Sustainable Development Indicators (companion material to ESDinds Deliverable 11)*. Brighton, UK: University of Brighton.

ESDinds. (2010d). Face-to-face interview with 'Maria' (pseudonym), project manager for organisation 'DB'. In G. Burford (Ed.), *ESDinds working paper: Transcripts of anonymised semi-structured interviews with participants in field tests of Set 1 Sustainable Development Indicators (companion material to ESDinds Deliverable 11)*. Brighton, UK: University of Brighton.

ESDinds. (2010e). Skype text interview with 'Ibrahim' (pseudonym), project manager for organisation 'TMSD'. In G. Burford (Ed.), *ESDinds working paper: Transcripts of anonymised semi-structured interviews with participants in field tests of Set 1 Sustainable Development Indicators (companion material to ESDinds Deliverable 11)*. Brighton, UK: University of Brighton.

ESDinds. (2010f). Skype verbal interview with 'Ingrid', project manager of organisation 'QU'. In G. Burford (Ed.), *ESDinds working paper: Transcripts of anonymised semi-structured interviews with participants in field tests of Set 1 Sustainable Development Indicators (companion material to ESDinds Deliverable 11)*. Brighton, UK: University of Brighton.

ESDinds. (2010g). Skype verbal interview with 'Rachel' (pseudonym), project officer of organisation 'JGSD'. In G. Burford (Ed.), *ESDinds working paper: Transcripts of anonymised semi-structured interviews with participants in field tests of Set 1 Sustainable Development Indicators (companion material to ESDinds Deliverable 11)*. Brighton, UK: University of Brighton.

ESDinds. (2011). *ESDinds: The development of values-based indicators and assessment tools for civil society organizations promoting education for sustainable development. Deliverable 17: final project report to European Commission Seventh Framework Programme (FP7/2007-2013): www.esdinds.eu*. Retrieved from Brighton

European Commission. (2007). *Seventh Framewok Programme guide for applicants: Cooperation: Theme 6: Environment (including climate change). Research for the benefit of specific groups: civil society organisations (BSG-CSO). FP7-ENV-2007-1. *. Retrieved from Brussels:

European Commission: EUROSTAT. *Sustainable development indicators and national accounts*. Retrieved from http://www.insee.fr/en/insee-statistique-publique/colloques/acn/pdf10/ravets.pdf

Farrell, R., & Hooker, C. (2013). Design, science and wicked problems. *Design Studies, 34*, 681-705.

Fehling, M., Nelson, B. D., & Venkatapuram, S. (2013). Limitations of the Millennium Development Goals: a literature review. *Global Public Health, 8*(10), 1109-1122. doi:10.1080/17441692.2013.845676

Feilzer, M., Y,. (2010). Doing mixed methods pragmatically: implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research, 4*(1), 6-16.

Fereday, J., & Muir-Cochrane, E. (2006). *Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development* (Vol. 5).

Ferguson, J. (1994). *The Anti-Politics Machine: "Development", Depoliticization and Bureaucratic Power in Lesotho*. Minneapolis, MN: University of Minnesota Press.

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology and Marketing, 17*(2), 105-120.

Foley, D. (2003). Indigenous epistemology and Indigenous Standpoint Theory. *Social Alternatives, 22*(1), 44-52.

Fontaine, J. R. J., Poortinga, Y. H., Delbeke, L., & Schwartz, S. H. (2008). Structural equivalence of the values domain across cultures: Distinguishing sampling fluctuations from meaningful variation. *Journal of Cross-Cultural Psychology, 39*, 345-365.

Foote, J., Gaffney, N., & Evans, J. R. (2010). Corporate social responsibility: implications for performance excellence. *Total Quality Management, 21*(8), 799-812.

Forlizzi, J. (2014). Research through design: a method for interaction design research in the field of human-computer interaction. *Design for All, 9*(4), 24-25.

Foucault, M. (1978). *Security, territory, population: Lectures at the College de France, 1977–1978. Ed. M. Senellart. Trans. G. Burchell.* . New York: Palgrave Macmillan.

Fraser, E. D. G., Dougill, A. J., Mabee, W. E., Reed, M. S., & McAlpine, P. (2006). Bottom up and top down: Analysis of participatory processes for sustainability indicator identification as a pathway to community empowerment and sustainable environmental management. *Journal of Environmental Management, 78*(2), 114-127.

Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers, 1*(1), 1-5.

Freire, P. (1970). *Pedagogy of the oppressed*. New York: The Seabury Press.

Freire, P. (1974). *Education for critical consciousness*. London: Sheed and Ward.

Fuad-Luke, A. (2007). Re-defining the purpose of (sustainable) design: enter the design enablers, catalysts in co-design. In J. Chapman & N. Gant (Eds.), *Designers, visionaries and other stories: a collection of sustainable design essays* (pp. 18-55). London: Earthscan.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385-400. doi:10.1016/0191-8869(86)90014-0

Garcia, S. M., Bazerman, M. H., Kopelman, S., Tor, A., & Miller, D. T. (2010). The price of equality: suboptimal resource allocations across social categories. *Business Ethics Quarterly, 20*(1), 75-88.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge*. London, Thousand Oaks & New Delhi: SAGE.

Goleman, D. (1998). *Working with emotional intelligence*. London: Bloomsbury.

Grainger, A. (2012). Forest sustainability indicator systems as procedural policy tools in global environmental governance. *Global Environmental Change, 22*, 147-160.

Gruys, M. L., Stewart, S. M., Goodstein, J., Bing, M. N., & Wicks, A. C. (2008). Values enactment in organizations: A multi-level examination. *Journal of Management, 34*(4), 806-843. doi:10.1177/0149206308318610

Gudmundsson, H. (2003). The policy use of environmental indicators - learning from evaluation research. *Journal of Transdisciplinary Environmental Studies 2*(2), 1-12.

Hák, T., Janoušková, S., & Moldan, B. (2016). Sustainable Development Goals: A need for relevant indicators. *Ecological Indicators, 60*, 565-573. doi:http://dx.doi.org/10.1016/j.ecolind.2015.08.003

Hak, T., Moldan, B., & Dahl, A. (2007). *Sustainability indicators: a scientific assessment. SCOPE (Scientific Committee on the Problems of Environment based in Paris), Vol. 67*. Washington, D.C.: Island Press.

Harder, M. K., & Burford, G. (2019). *Measuring intangible values: rethinking how to evaluate socially beneficial actions*. Abingdon, UK: Routledge.

Harder, M. K., Burford, G., & Hoover, E. (2013). What is participation? Design leads the way to a cross-disciplinary framework. *Design Issues, 29*(4), 41-57.

Harder, M. K., Burford, G., & Hoover, E. (2014). *From 'sustainable production' to 'production as sustainability': the emergence of values-focused evaluation.* Paper presented at the Global Research Forum on Sustainable Production and Consumption, Shanghai, China.

Harder, M. K., Velasco, I., Burford, G., Podger, D., Janouskova, S., Piggot, G., & Hoover, E. (2014). Reconceptualizing 'efffectiveness' in environmental projects: can we measure values-related achievements? *Journal of Environmental Management, 139*, 120-134.

Hawkes, J. (2001). *The fourth pillar of sustainability: culture's essential role in public planning*. Victoria: Common Ground Publishing Pty Ltd in association with the Cultural Development Network (Vic).

Hedlund-de Witt, A. (2011). The rising culture and worldview of contemporary spirituality: A sociological study of potentials and pitfalls for sustainable development. *Ecological Economics, 70*, 1057-1065.

Heimlich, J. E. (2010). Environmental education evaluation: Reinterpreting education as a strategy for meeting mission. *Evaluation and Program Planning, 33*, 180-185.

Henshaw, J. M. (2006). Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth. In *Measurement in business: what gets measured gets done* (pp. 55-65). Baltimore, MD: Johns Hopkins University Press.

Heugens, P. P. M. A. R., & Scherer, A. G. (2010). When organization theory met business ethics: toward further symbioses. *Business Ethics Quarterly, 20*(4), 643-672.

Hewlin, P. F. (2003). And the award for best actor goes to ... : Facades of conformity in organizational settings. *Academy of Management Review, 28*(4), 633-642.

Hinkel, J. (2011). ''Indicators of vulnerability and adaptive capacity'': Towards a clarification of the science–policy interface. . *Global Environmental Change, 21*, 198-208.

Hitlin, S., & Piliavin, J. A. (2004). Values: Reviving a dormant concept. *Annual Review of Sociology, 30*(1), 359-393.

Hochachka, G. (2005). *Developing sustainability, developing the self: an integral approach to individual and community development*. Victoria, Canada: Polis Project.

Holland, D., & Albrecht, C. (2013). The worldwide academic field of business ethics: scholars' perceptions of the most important issues. *Journal of Business Ethics, 117*(4), 777-788.

Hoover, E., Burford, G., Dredge, L., & Harder, M. K. (2015). *Starting from values: evaluating intangible legacies*. Brighton, UK: University of Brighton.

Hopwood, W., Mellor, M., & O'Brien, G. (2005). Sustainable development: Mapping different approaches. *Sustainable Development, 13*, 38-52.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications.

Huhtala, M., Feldt, T., Lämsä, A.-M., Mauno, S., & Kinnunen, U. (2011). Does the ethical culture of organisations promote managers' occupational well-being? Investigating indirect links via ethical strain. *Journal of Business Ethics, 101*(2), 231-247. doi:10.1007/s10551-010-0719-3

Hulme, M. (2010). Problems with making and governing global kinds of knowledge. *Global Environmental Change, 20*, 558-564.

IAEG-SDG. (2015a). Inter-Agency and Expert Group on Sustainable Development Goal Indicators: IAEG-SDG Open Consultation on Grey Indicators. Retrieved from http://unstats.un.org/sdgs/iaeg-sdgs/open-consultation-3

IAEG-SDG. (2015b). Inter-Agency and Expert Group on Sustainable Development Goal Indicators: Open Consultation on Green Indicators. Retrieved from http://unstats.un.org/sdgs/iaeg-sdgs/open-consultation-2

IAEG-SDG. (2016a). Inter-Agency and Expert Group on Sustainable Development Goal Indicators: IAEG-SDG membership.

IAEG-SDG. (2016b). *Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators to the 47th Session of the United Nations Statistical Commission, *. Retrieved from New York:

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American sociological review, 65*(1), 19-51.

Inglehart, R., & Welzel, C. (2005). *Modernization, Cultural Change and Democracy*. New York: Cambridge University Press.

Interreligious Statement Towards Rio+20. (2012). *Towards Rio+20 and Beyond - A Turning Point in Earth's History*. Retrieved from Kalk Bay, South Africa: http://safcei.org/wp-content/uploads/2012/04/Interreligious-Statement-Towards-Rio1.pdf

Ioppolo, G., Cucurachi, S., Salomone, R., Saija, G., & Shi, L. (2016). Sustainable local development and environmental governance: a strategic planning experience. *Sustainability, 8*, 180.

Jackson, T. J. (2005). *A report to Sustainable Development Research Network. Motivating Sustainable Consumption: a review of evidence on consumer behavior and behavioral change*. Retrieved from

Jickling, B. (2013). Normalizing catastrophe: an educational response. *Environmental Education Research, 19*(2), 161-176. doi:10.1080/13504622.2012.721114

Jickling, B., & Wals, A. E. J. (2008). Globalization and environmental education: looking beyond sustainable development. *Journal of Curriculum Studies, 40*(1), 1-21.

Kagan, S. (2011). *Art and sustainability: connecting patterns for a culture of complexity*. Bielefeld, Germany: Transcript Verlag.

Kemaghan, K. (2003). Integrating Values into Public Service: The Values Statement as Centerpiece. *Public Administration Review, 63*(6), 711-719.

Kennedy, E. H., Beckley, T. M., McFarlane, B. L., & Nadeau, S. (2009). Why we don't 'walk the talk': Understanding the environmental values/behaviour gap in Canada. . *Research in Human Ecology, 16*, 151-160.

Kluckhohn, C. (1951). Values and value-orientations in the theory of action. In T. Parsons & E. Shils (Eds.), *Toward a general theory of action* (pp. 388-433). Cambridge, MA: Harvard University Press.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? . *Environmental Education Research, 8*(3), 239-260.

Kopelman, R. E., Rovenpor, J. L., & Guan, M. (2003). The Study of Values: Construction of the fourth edition. . *Journal of Vocational Behavior, 62*, 203-220.

Koroneos, C. J., & Rokos, D. (2012). Sustainable and integrated development - a critical analysis. *Sustainability, 4*, 141-153.

Krank, S., Wallbaum, H., & Grêt-Regamey, A. (2013). Perceived contribution of indicator systems to sustainable development in developing countries. *Sustainable Development, 21*(1), 18-29. doi:10.1002/sd.496

Lang, D. J. (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science, 7*, 25-43.

Lather, P. (1986). Issues of validity in openly ideological research: between a rock and a soft place. *Interchange, 17*(4), 63-84.

Leach, M., Scoones, I., & Stirling, A. (2010). *Dynamic sustainabilities: technology, environment, social justice*. London: Earthscan.

Lele, S., & Norgaard, R. B. (1996). Sustainability and the scientist's burden. *Conservation Biology, 10*, 354-365.

Lencioni, P. M. (2002). Make your values mean something. *Harvard Business Review, 80*(7), 113-117.

Littig, B., & Griessler, E. (2005). Social sustainability: a catchword between political pragmatism and social theory. *International Journal of Sustainable Development, 8*(1-2), 65-79.

Lomazzi, M., Borisch, B., & Laaser, U. (2014). The Millennium Development Goals: experiences, achievements and what's next. *Global health action, 7*, 23695-23695. doi:10.3402/gha.v7.23695

Lotz-Sisitka, H., Ali, M. B., Mphepo, G., Chaves, M., Macintyre, T., Pesanayi, T., . . . McGarry, D. (2016). Co-designing research on transgressive learning in times of climate change. *Current Opinion in Environmental Sustainability, 20*, 50-55.

Lotz-Sisitka, H., Wals, A. E. J., Kronlid, D., & McGarry, D. (2015). Transformative, transgressive social learning: rethinking higher education pedagogy in times of systemic global dysfunction. *Current Opinion in Environmental Sustainability, 16*, 73-80.

Lozano, R., Llobet, J., & Tideswell, G. (2016). *The process of assessing and reporting sustainability at universities: Preparing the report of the University of Leeds*. Retrieved from Barcelona:

Lu, Y., Nakicenovic, N., Visbeck, M., & Stevance, S. (2015). Five priorities for the UN Sustainable Development Goals. *Nature, 520*, 432-433.

Machicado, C. A., & Davis, H. (1988). A Values Assessment Application Among Bolivian Managers. *The Journal of Social Psychology, 128*(5), 691-693.

Maio, G. R., Olson, J. M., Allen, L., & Bernard, M. M. (2001). Addressing discrepancies between values and behavior: The motivating effect of reasons. *Journal of Experimental Social Psychology, 37*(2), 104-117.

Manning, R. (2009). *Using indicators to encourage development: lessons from the Millennium Development Goals*. Retrieved from Copenhagen:

Marks, N., Abdallah, S.,Simms, A., and Thompson, S. (2006). *The (Un)happy Planet Index : An Index of Human Well-Being and Environmental Impact* London: New Economics Foundation.

Maturana, H., & Varela, F. J. (1991). *Autopoiesis and cognition: The realization of the living*. Dordrecht, The Netherlands: D. Reidel Publishing Company.

McCool, S., & Stankey, G. (2004). Indicators of sustainability: challenges and opportunities at the interface of science and policy. *Environmental Management, 33*(3), 294-305.

McFerran, B., Acquino, K., & Duffy, M. (2010). How personality and moral identity relate to individuals' ethical ideology. *Business Ethics Quarterly, 20*(1), 35-56.

McGregor, S. L. T., & Volckmann, R. (2013). Transversity: transdisciplinarity in higher education. In G. Hampson & M. Rich-Tolsma (Eds.), *Leading transformative higher education* (pp. 58-81). Olomouc, Czech Republic: Palacky University Press.

McNiff, J. (2013). *Action research: principles and practice*. London Routledge.

Meadows, D. (1998). *Indicators and Information Systems for Sustainable Development: Report to the Balaton Group*. Retrieved from Hartland Four Corners:

Meglino, B. M., & Ravlin, E. C. (1998). Individual values in organizations: Concepts, controversies, and research. *Journal of Management, 24*(3), 351-389.

Mertens, D. (2010). *Transformative Mixed Methods Research* (Vol. 16).

Mertens, D. M. (2011). Mixed Methods as Tools for Social Change. *Journal of Mixed Methods Research, 5*(3), 195-197. doi:10.1177/1558689811418095

Meyer, J. H. F., & Shanahan, M. P. (2004). Developing metalearning capacity in students — Actionable theory and practical lessons learned in first-year economics. *Innovations in Education and Teaching International, 41*(4), 443-458.

Murga-Menoyo, M. (2014). Learning for a sustainable economy: Teaching of green competencies in the university. *Sustainability, 6*(5), 2974-2992.

Naess, A. (1973). The shallow and the deep, long-range ecology movements. *Inquiry, 16*, 95-100.

Nicolescu, B. (1998). The transdisciplinary evolution of the university: condition for sustainable development. *Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires, 12*. http://ciret-transdisciplinarity.org/bulletin/b12c8.php

Nicolescu, B. (2002). *Manifesto of transdisciplinarity*. New York: State Univesity of New York (SUNY) Press.

Nicolescu, B. (2006). Transdisciplinarity - past, present and future. In B. Haverkort & C. Reijntjes (Eds.), *Moving Worldviews - Reshaping sciences, policies and practices for endogenous sustainable development* (pp. 142-166). Amsterdam, The Netherlands: COMPAS Editions.

Nicolescu, B. (2010). Methlology of transdisciplinarity - levels of reality, logic of the included middle and complexity. *Transdisciplinary Journal of Engineering and Science, 1*(1), 19-38.

Nurse, K. (2006). *Culture as the Fourth Pillar of Sustainable Development*. Retrieved from London, UK:

Nygreen, K. (2017). Troubling the discourse of both/and: Technologies of neoliberal governance in community-based educational spaces. *Policy Futures in Education, 15*(2), 202-220. doi:10.1177/1478210317705739

O'Donoghue, T., & Chapman, E. (2010). Problems and prospects in competencies-based education: a curriculum studies perspective. *Education Research and Perspectives, 37*(1), 85-104.

Ofsted. (2018). *Ofsted Annual Report 2017/18*. Retrieved from London:

Oladottir, H. (2014). *UNESCO Education for Sustainable Development 2005-2014*. Retrieved from http://menntuntilsjalfbaerni.weebly.com/uploads/6/2/6/2/6262718/unesco_5_pillars_for_esd.pdf

Onwuegbuzie, A., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools, 13*(1), 48-63.

Onwuegbuzie, A., & Leech, N. L. (2005). On becoming a pragmatic researcher: the importance of combining qualitative and quantitative research methodologies. *International Journal of Social Research Methodology, 8*(5), 375-387.

Onwuegbuzie, A., Witcher, A., Collins, K., Filer, J., Wiedmaier, C. l., & Moore, C. (2007). Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-Methods Analysis. *American Educational Research Journal, 44*(1), 113-160. doi:10.3102/0002831206298169

Ostrom, E. (2010). Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change, 20*, 550-557.

Oxford English Dictionary. (2013). *Definition of 'values'*. In. Retrieved from http://oxforddictionaries.com/definition/english/value

Palmer, M., & Wagner, K. (2012). *ValuesQuest: the search for values which will make a world of difference*. Retrieved from Zurich, Switzerland:

Patten, T. (2013). *Enacting an integral revolution: how can we have truly radical conversations at a time of global crisis?* Paper presented at the Integral Theory Conference.

Patton, M. Q. (2007). Process use as a usefulism. *New Directions for Evaluation, 2007*(116), 99-112. doi:10.1002/ev.246

Pedgley, O. (2007). Capturing own design activity. *Design Studies, 28*, 463-483.

Peng, K. P., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods, 2*(4), 329-344.

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Assessment of character strengths. In G. P. Koocher, J. C. Norcross, & S. S. I. Hill (Eds.), *Psychologists' desk reference (2nd ed)* (pp. 93-98). New York: Oxford University Press.

Pfahl, S. (2005). Institutional sustainability. *International Journal of Sustainable Development, 8*(1-2), 80-96.

Plotkin, B. (2009). *Nature and the human soul: cultivating wholeness and community in a fragmented world*. Novato, CA: New World Library.

Podger, D., Hoover, E., Burford, G., Hak, T., & Harder, M. K. (2016). Revealing values in a complex environmental program: a scaling up of values-based indicators. *Journal of Cleaner Production, 134*, 225-238.

Podger, D., Piggot, G., Zahradnik, M., Janouskova, S., Velasco, I., Hak, T., . . . Harder, M. K. (2010). The Earth Charter and the ESDinds initiative: developing indicators and assessment tools for civil society organisations to examine the values dimensions of sustainability projects. *Journal of Education for Sustainable Development, 4*(2), 297-305.

Podger, D., Velasco, I., Amezcua Luna, C., Burford, G., & Harder, M. K. (2013). Can values be measured? Significant contributions from a small civil society organisation through action research evaluation. *Action Research, 11*(1), 8-30.

Polanyi, M. (2009). *The tacit dimension* (Revised edition ed.). Chicago: University of Chicago Press.

Pryshlakivsky, J., & Searcy, C. (2012). Sustainable development as a wicked problem. In S. F. Kovacic & A. Sousa-Poza (Eds.), *Managing and engineering in complex situations* (pp. 109-128). Dordrecht: Springer Netherlands.

Rametsteiner, E., Puelzl, H., Alkan-Olsson, J., & Frederiksen, P. (2011). Sustainability indicator development: Science or political negotiation? *Ecological Indicators, 11*(1), 61-70.

Raths, L. E., Harmin, M., & Simon, S. B. (1978). *Values and teaching* (2nd ed.). Columbus, OH, USA: C. E. Merrill.

Raworth, K. (2017). *Doughnut economics: seven ways to think like a 21st century economist*. White River Junction, Vermont: Chelsea Green Publishing.

Redclift, M., & Benton, T. (1994). Introduction. In *Social Theory and the Global Environment*. London: Routledge.

Reed, M. S., & Dougill, A. J. (2002). Participatory selection process for indicators of rangeland condition in the Kalahari. *Geographical Journal, 168*, 224-234. doi:10.1111/1475-4959.00050

Reed, M. S., Dougill, A. J., & Baker, T. R. (2008). Participatory indicator development: What can ecologists and local communities learn from each other? *Ecological Applications, 18*(5), 1253-1269.

Reed, M. S., Fraser, E. D. G., & Dougill, A. J. (2006). An adaptive learning process for developing and applying sustainability indicators with local communities. *Ecological Economics, 59*(4), 406-418.

Reed, M. S., Fraser, E. D. G., Morse, S., & Dougill, A. J. (2005). Integrating Methods for Developing Sustainability Indicators to Facilitate Learning and Action. *Ecology & Society, 10*(1), 1-6.

Reino, A., & Vadi, M. (2010). *What factors predict the values of an organization and how?* Faculty of Economics & Business Administration Working Paper. University of Tartu, Estonia.

Rescher, N. (1982). *Introduction to Value Theory*. Lanham, MD: University Press of America.

Research Department of the Universal House of Justice. (1993). Homosexuality.

Reynolds, S. J. (2008). Moral attentiveness: who pays attention to the moral aspects of life? *Journal of Applied Psychology, 93*, 1027-1041.

Ribeiro, M. M., Hoover, E., Burford, G., Buchebner, J., Lindenthal, T., & Leal Filho, W. (2016). Values: a bridge between sustainability and institutional assessment-a case study from BOKU University. *International Journal of Sustainability in Higher Education, 17*(1).

Robinson, J. (2004). Squaring the circle? Some thoughts on the idea of sustainable development. *Ecological Economics, 48*, 369-384.

Rokeach, M. (1973). *The Nature of Human Values*. New York: The Free Press.

Rokeach, M. (1979a). From individual to institutional values: With special reference to the values of science. In *Understanding human values* (pp. 47-70). New York: Free Press.

Rokeach, M. (1979b). *Understanding human values*. New York: Free Press.

Rokeach, M. (1985). inducing change and stability in belief systems and personality structures. *Journal of Social Issues, 41*(1), 153-171.

Rosenström, U. (2006). Exploring the policy use of sustainable development indicators: interviews with Finnish politicians. *The Journal of Transdisciplinary Environmental Studies, 5*(1-2), 1-13.

Rosenström, U., & Kyllönen, S. (2007). Impacts of a participatory approach to developing national level sustainable development indicators in Finland. *Journal of Environmental Management, 84*, 282-298.

Rychen, D. S., & Salganik, L. H. (2001). *Defining and selecting key competencies*. Ashland, OH, USA: Hogrefe & Huber Publishers.

Sabo Flores, K. (2008). *Youth participatory evaluation*. San Francisco, CA, USA: Jossey-Bass.

Saith, A. (2006). From universal values to Millennium Development Goals: lost in translation. *Development and Change, 37*(6), 1167-1199.

Sanders, E. B.-N., & Stappers, P. J. (2012). *Convivial toolbox: generative research for the front end of design*. Amsterdam, The Netherlands: BIS Publishers.

Schlater, J. D. (1969). Investigating values underlying family decisions. *Research Bulletin 23*.

Schlater, J. D., & Sontag, M. (1994). Toward the measurement of human values. *Family and Consumer Science Research Journal, 23*(1), 4-25.

Schwartz, S. H. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1-65.

Schwartz, S. H. (1994). Are there universal aspects in the structure and content of human values? *Journal of Social Issues, 50*(4), 19-45.

Schwartz, S. H. (2007). *Basic Human Values: Theory, Methods, and Applications*. The Hebrew University of Jerusalem. Retrieved from http://segr-did2.fmag.unict.it/Allegati/convegno%207-8-10-05/Schwartzpaper.pdf

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology, 53*(3), 550-562.

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values - extensions and cross-cultural replications. *Journal of Personality and Social Psychology, 58*(5), 878-891.

Schwartz, S. H., & Bilsky, W. (1990). Toward a Theory of the Universal Content and Structure of Values: Extensions and Cross-Cultural Replications. *Journai of Personality and Social Psychology, 58*(5), 878-889.

Schwartz, S. H., Lehmann, A., & Roccas, S. (1999). Multimethod probes of basic human values. In J. Adamopoulos & Y. Kashima (Eds.), *Social psychology and culture context: Essays in honor of Harry C. Triandis* (pp. 107–123). Newbury Park, CA: Sage.

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology, 32*(5).

Seamon, D., & Zajonc, A. (1998). *Goethe's way of science: a phenomenology of nature*. Albany, NY: SUNY.

Sleurs, W. (2008). *Competencies for ESD (Education for Sustainable Development) teachers: a framework to integrate ESD in the curriculum of teacher training institutes*. Retrieved from Brussels:

Smyth, J. (1999). Is there a future for education consistent with Agenda 21? *Canadian Journal of Environmental Education, 4*, 69-82.

Soucek, V. (1993). Flexible education and new standards of communication competence. In P. Watkins, V. Soucek, S. Robertson, & P. Brown (Eds.), *Economising Education: The Post-Fordist Directions* (pp. 33-72). Victoria: Deakin University.

Spangenberg, J. H. (2002). Institutional sustainability indicators: an analysis of the institutions in Agenda 21 and a draft set of indicators for monitoring their effectivity. *Sustainable Development, 10*(2), 103-115. doi:10.1002/sd.184

Spangenberg, J. H., Pfahl, S., & Deller, K. (2002). Towards indicators for institutional sustainability: lessons from an analysis of Agenda 21. *Ecological Indicators, 2*(1–2), 61-77. doi:http://dx.doi.org/10.1016/S1470-160X(02)00050-X

Spencer, C., & Blades, M. (2006). *Children and their environments: Learning, using and designing spaces*: Cambridge University Press.

Spranger, E. (1928). *Types of men [Originally published in German by Niemeyer, Halle (Saale), 1914, as Lebensformen; translated by P. J. W. Pigors]*. New York: G. E. Stechert Company.

Springett, J. (2001). Participatory approaches to evaluation in health promotion In M. G. I. Rootman, B. Hyndman, D. V. McQueen, L. Potvin, J. Springett, & E. Ziglio (Ed.), *Evaluation in health*

promotion: Principles and perspectives (pp. 83-105). Copenhagen: WHO Regional Publications.

Staff, A. I. o. A. K. R. (2009). *AIA Research Primer*. Retrieved from Washington, D.C., USA:

Stapleton, L. M., & Garrod, G. (2008). Policy preceding possibility? Examining headline composite sustainability indicators in the United Kingdom *Social Indicators Research, 87*, 495-502.

Sterling, S. (2001). *Sustainable education: revisioning learning and change*. Cambridge, UK: Green Books.

Sterling, S. (2004). Higher education, sustainability and the role of systemic learning. In P. B. Corcoran & A. E. J. Wals (Eds.), *Higher education and the challenge of sustainability: problematics, promise, and practice* (pp. 49-70). Dordrecht: Kluwer Academic Publishers.

Suspitsyna, T. (2010). *Accountability in American education as a rhetoric and a technology of governmentality* (Vol. 25).

Swyngedouw, E. (2004). Globalisation or 'glocalisation'? Networks, territories and rescaling. *Cambridge Review of International Affairs, 17*(1), 25-48.

Talamo, A., & Pozzi, S. (2011). The tension between dialogicality and interobjectivity in cooperative activities. *Culture & Psychology 17*(3), 302-318.

Tamas, P. (in press). Systematic review of methods in low-consensus fields: supporting commensuration through `construct-centered methods aggregation' in the case of climate change vulnerability research. *PLOS One*.

Tawil, S., & Cougoureux, M. (2013). *Revisiting 'Learning: The Treasure Within': Assessing the influence of the 1996 Delors Report*. Retrieved from Paris, France:

The Australian Institute of Architects. (2009). *Understanding research excellence in architecture*. Retrieved from Barton, Australia:

Thompson, S., & Barton, M. (1994). *Ecocentric and Anthropocentric Attitudes Toward the Environment* (Vol. 14).

Till, J. (2008). *Architectural research: three myths and one model*. Retrieved from London, UK:

Torbert, W. (2001). The practice of action inquiry. In P. Reason & H. Bradbury (Eds.), *Handbook of action research* (pp. 250-260). London: Sage.

Torpe, L., & Lolle, H. (2010). Identifying Social Trust in Cross-Country Analysis: Do We Really Measure the Same? *Social Indicators Research Online First™*.

Tufford, L., & Newman, P. (2010). Bracketing in qualitative research. *Qualitative Social Work, 11*(1), 80-96.

UN-PFII. (2006). *Report of the Meeting on Indigenous Peoples and Indicators of Well-Being, Ottawa, 22-23 March 2006*. Retrieved from New York, NY:

UN Department of Economic and Social Affairs, D. f. S. D. (1999). *Testing the CSD Indicators of Sustainable Development, Interim Analysis: Testing Process, Indicators and Methodology Sheets. Technical Paper.* New York, NY: UNDESA.

UN Division for Sustainable Development. Department of Policy Co-ordination and Sustainable Development. (1995). *Work Programme on Indicators of Sustainable Development, Document UN/E/CN.17/1995/18*. Retrieved from New York:

UN Division for Sustainable Development. Department of Policy Co-ordination and Sustainable Development. (1996). *Indicators of sustainable development, framework and methodologies*. Retrieved from New York:

UN Division for Sustainable Development. UN Department of Economic and Social Affairs. (2000). *Indicators of sustainable Development, framework and core set*. Retrieved from New York:

UNECE. (2011). *Learning from Each Other: Achievements, Challenges and Ways Forward: Second Evaluation Report of the Implementation of the UNECE ESD Strategy*. Retrieved from Geneva, Switzerland:

UNESCO. (1982). *Mexico City Declaration on Cultural Policies*. Retrieved from Paris: http://portal.unesco.org/culture/en/files/35197/11919410061mexico_en.pdf/mexico_en.pdf

UNESCO. (2004). *International Implementation Scheme for the Decade of Education for Sustainable Development*. Retrieved from Paris:

UNESCO. (2012). *Shaping the Education of Tomorrow: 2012 Full-length Report on the UN Decade of Education for Sustainable Development*. Retrieved from Paris, France:

UNESCO. (2013). *Culture in the post-2015 sustainable development agenda: why culture is key to sustainable development. Background paper to the 2013 Hangzhou Congress on Culture and Development, 14-17 May 2013, Hangzhou, People's Republic of China.* http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/images/Post2015Sustainable DevelopmentAgendaENG.pdf

UNESCO. (2014). *Aichi-Nagoya Declaration on Education for Sustainable Development* Retrieved from Paris, France:

UNESCO. (2016). What is Education for Sustainable Development? Retrieved from http://en.unesco.org/themes/education-sustainable-development/what-is-esd

UNESCO. (2017). *Accountability in education: meeting our commitments - Global Education Monitoring Report 2017/18*. Retrieved from Paris:

United Nations. (2002). *Report of the World Summit on Sustainable Development, Johannesburg, South Africa, 26 August - 4 September 2002*. Retrieved from New York:

United Nations. (2012). *The future we want: Outcome document of the United Nations Conference on Sustainable Development in Rio de Janeiro from 20-22 June 2012*. Retrieved from New York:

United Nations. (2015a). *The Millennium Development Goals Report, 2015*. Retrieved from New York:

United Nations. (2015b). *Transforming our world: the 2030 agenda for sustainable development*. Retrieved from New York:

United Nations Development Group. (2013). *A million voices: the world we want - A sustainable future with dignity for all*. Retrieved from New York:

United Nations Economic Commission for Europe (UNECE). (2011). *Learning for the future: Competences in Education for Sustainable Development. United Nations Economic Commission for Europe Strategy for Education for Sustainable Development, ECE/CEP/AC.13/2011/6*. Retrieved from Utrecht, The Netherlands:

United Nations General Assembly. (2000). *United Nations Millennium Declaration*. Retrieved from New York: http://www.un.org/millennium/declaration/ares552e.htm

United Nations Millennium Campaign. (2015). *We the peoples, celebrating 8 million voices: the MY World global survey for a better world*. Retrieved from New York:

van de Weijer, M., Van Cleempoel, K., & Heynen, H. (2014). Positioning research and design in academia and practice: a contribution to a continuing debate. *Design Issues, 30*(2), 17-29.

van der Hoeven, F. (2011). Mind the evaluation gap: reviewing the assessment of architectural research in the Netherlands. *Architectural Research Quarterly, 15*(2), 177-187.

Verquer, M. L., Beehr, T. A., & Wagner, S. H. (2003). A meta-analysis of relations between person-organization fit and work attitudes. *Journal of Vocational Behavior, 63*(3), 473-489.

Vladimirova, K., & Le Blanc, D. (2016). Exploring Links Between Education and Sustainable Development Goals Through the Lens of UN Flagship Reports. *24*(4), 254-271. doi:doi:10.1002/sd.1626

von Groddeck, V. (2011). Rethinking the Role of Value Communication in Business Corporations from a Sociological Perspective - Why Organisations Need Value-Based Semantics to Cope with Societal and Organisational Fuzziness. *Journal of Business Ethics, 100*(1), 69-84. doi:10.1007/s10551-011-0769-1

Vucetich, J. A., & Nelson, M. P. (2010). Sustainability: virtuous or vulgar? *BioScience, 60*(7), 539-544.

Waage, J., Banerji, R., Campbell, O., Chirwa, E., Callendar, G., Dieltiens, V. D., A., . . . Kingdon, G. (2010). The Millennium Development Goals: a cross-sectoral analysis and principles for goal setting after 2015. *The Lancet, 376*(9745), 991-1023.

Waas, T., Hugé, J., Verbruggen, A., & Wright, T. (2011). Sustainable development: a bird's eye view. *Sustainability, 3*, 1637-1661.

Walker, S. (2006). *Sustainable by Design: Explorations in Theory and Practice*. London, UK: Earthscan / James and James.

Walker, S. (2007). Design redux. In J. Chapman & N. Gant (Eds.), *Designers, visionaries and other stories* (pp. 56-75). London: Earthscan.

Walker, S., & Giard, J. (2013). General introduction: design for sustainability - a reflection. In S. Walker & J. Giard (Eds.), *Handbook of design for sustainability* (pp. 1-12). London & New York: Bloomsbury Academic.

Wals, A. E., & Corcoran, P. B. (2006). 14. Sustainability as an Outcome of Transformative Learning. *Drivers and barriers for implementing sustainable development in higher education*, 103.

Wals, A. E. J., & Jickling, B. (2002). 'Sustainability' in higher education: from doublethink and newspeak to critical thinking and meaningful learning. *International Journal of Sustainability in Higher Education, 3*(3), 221-232.

Warren, D. E., Gaspar, J. P., & Laufer, W. F. (2014). Is formal ethics training merely cosmetic? A study of ethics training and ethical organizational culture. *Business Ethics Quarterly, 24*(1), 85-117.

Wiek, A., Withycombe, L., & Redman, C. L. (2011). Key competencies in sustainability: a reference framework for academic program development. *Sustainability Science, 6*, 203-218.

Wilber, K. (2004). The integral vision of healing. In M. Schlitz, T. Amorok, & M. Micozzi (Eds.), *Consciousness and healing: integral approaches to mind-body medicine*. St Louis, MO: C. V. Mosby.

Williams, R. M. (1979). Changes and stability in values and value systems, a sociological perspective. In M. Rokeach (Ed.), *Understanding human values* (pp. 15-46). New York: The Free Press.

Wood, J. (2007). Relative abundance: Fuller's discovery that the glass is always half full. In J. Chapman & N. Gant (Eds.), *Designers, visionaries and other stories: a collection of sustainable design essays* (pp. 96-115). London: Earthscan.

Woodley, E., Crowley, E., Dookie, C., Carmen, E. (2006). *Cultural indicators of Indigenous Peoples' food and agro-ecological systems*. Retrieved from Rome:

World Commission on Environment and Development. (1987). Chapter 12. Towards Common Action: Proposals for Institutional and Legal Change. In Our Common Future: Report of the World Commission on Environment and Development. Geneva: United Nations. Retrieved from http://www.un-documents.net/ocf-12.htm.

Zimmerman, J., & Forlizzi, J. (2008). The role of design artifacts in design theory construction. *Artifact, 2*(1), 41-45.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). *Research through design as a method for interaction design in human-computer interaction.* Paper presented at the Conference on Human Factors in Computing Systems, New York.

Summary

Evaluating the success or failure of educational initiatives, whether at the level of individual students, teachers and institutions or at national and global levels, relies on a shared understanding of what 'success' or 'failure' might look like in practice. This, in turn, cannot be defined in an objective way because it depends on the values, mindsets and priorities of diverse stakeholders - which may be very different from one another, or even mutually conflicting. In order to design assessment tools to 'measure what we treasure', as called for in the 2015 Millennium Development Report, we must first embark on the challenging task of creating appropriate indicators. Only then can data be collected and analysed, conclusions drawn, and policy priorities reviewed.

The majority of educational indicators and evaluation systems are designed from the top down, and based on the priorities of a single stakeholder – usually a government or a donor organisation. This has perpetuated the use of indicators such as Gross Domestic Product (GDP), which effectively 'reward' governments for engaging in activities that are environmentally and socially destructive. Within the field of education for sustainability, the design of evaluations and assessments tends to be based on the accumulation of a large number of poorly-defined indicators, which are often difficult to measure.

In this thesis, I have described four separate research studies conducted as part of the ESDinds project ('Creating Indicators and Assessment Tools for Civil Society Organizations Promoting Education for Sustainable Development'), which was led by a consortium of two universities and four civil society organizations from 2008-2011. The project aimed to explore *inductive* ways of developing indicators in the context of non-formal education for sustainability – drawing them out from participants' comments about what they find valuable, meaningful and worthwhile within a shared context of practical action, rather than deducing a framework from theory. These approaches were also *intersubjective,* in the sense of seeking to understand and empathise with each other's perspectives, identify a common core if applicable, and build consensus about how to represent what matters to the group as a whole within the specified context. The indicators created in this way can be described as 'values-based'. A key aspect of the approach described here (as an illustrative example) is the creation of resource materials to challenge people's preconceived assumptions, promote dialogue, and raise awareness of marginalised groups and world-views.

In the first of the four studies presented here, I describe the organisational impacts of conducting field trials with a 'values-based evaluation' approach in eight different organisations. The observed impacts included an improved understanding of how values-related vocabulary can be

linked to specific actions, feelings and behaviours (which I have termed 'values conceptualization') and incorporation of this newly acquired vocabulary, with its referents, into training, PR and other conversations within and beyond the respective organizations ('values mainstreaming'). The data analysis also indicates an increased awareness of evaluation methodologies and knowledge of how to apply specific assessment methods, and changes in self-perception, perceptions of others, and professional relationships within the organizations. These preliminary findings are discussed in relation to wider academic discourse on values communication in organisations.

Having established the potential for organisational benefits of inductive / intersubjective evaluation design ('values-based evaluation') in practice, I proceed to investigate its theoretical validity in more depth. Specifically, I refute the theoretical argument which is often used to suggest that values enactment is impossible to measure, and replace it with a 'context-specific measurability' argument which suggests that intersubjective conceptualization of values can be achieved within clearly-defined practical contexts. This opens up new possibilities for the operationalization of what has been termed the 'missing pillar' of sustainability – the various intangible dimensions (cultural, aesthetic, political, spiritual, etc.) that are excluded from the traditional 'three-pillar' model comprising environmental, economic and social dimensions. The theoretical work is followed up with a short case study of the use of inductive/intersubjective approaches to evaluate an online course in sustainability leadership (Chapter 3), and subsequently a rigorous investigation of how such approaches might contribute to conversations around indicator design in relation to the Education for Sustainable Development (ESD) target of the Sustainable Development Goals. I demonstrate that the 'values-based' (inductive and intersubjective) approach can be complementary to traditional United Nations indicator development processes, and is effective in operationalising intangible dimensions of ESD such as 'appreciation of cultural diversity'.

As a contrast to the high-level policy dialogues surrounding the SDGs and their indicators, I then explore the applicability of inductive and intersubjective approaches to 'indicator' design at the level of individual schools or classes. I do not discuss them as evaluative interventions in this case, although they could be used as such, but as tools for stimulating reflection and learning about values and sustainability-related issues among students (and educators). I describe a process of 'Research through Design' to develop toolkits that can transform the way in which ESD / Education for Sustainable and Responsible Living (EfSRL) is approached in schools, from the provision of factual knowledge about environmental problems to a holistic strategy focusing on the development of competencies. I raise the possibility that this type of education could contribute to bridging the gaps between values, discourse and action, and discuss my findings in the light of recent explorations of transformative, transgressive and transdisciplinary (which I term 'triple-T') learning initiatives.

About the author

Ashley Jay Brockwell (formerly known as Gemma Burford) was born in Southampton, England on 5 November 1977, graduating from the University of Oxford with a MBiochem in Molecular and Cellular Biochemistry in June 1999 and gaining an MSc in Environmental Anthropology (with distinction) from the University of Kent in 2002. This marked the beginning of a ten-year research career with the Global Initiative for Traditional Systems of Health, working at the interface of global health policy, integrative medicine, ethnobotany, ethnopharmacology and anthropology – initially as a conference organiser, then as a research assistant, and finally as an international research associate. This period of research (some of it conducted in Tanzania and Kenya) yielded more than 20 peer-reviewed publications, including journal articles, book chapters, and two co-edited volumes on traditional, complementary and alternative medicine.

During a ten-year stay in Tanzania, Ash co-founded the non-governmental organization Aang Serian ('House of Peace'), which built a school in a Maasai village, developed an innovative intercultural curriculum that was recognised in a United Nations report on implementation of the Convention on Biological Diversity, created an Indigenous Media Project, and catalysed an arts-based grassroots campaign against female genital mutilation and forced child marriage.

After returning to the UK in 2010, Ash took up a position as Research Officer (later Research Fellow) in the Values and Sustainability Research Group at the University of Brighton. The research that constitutes this thesis was conducted between 2010 and 2016, along with additional studies leading to a further nine collaborative publications in peer-reviewed journals, a chapter in an edited volume, and the book *Measuring Intangible Values: Rethinking How To Evaluate Socially Beneficial Actions* (to be published by Routledge in 2019). During this time, Ash received the Community Animation and Social Innovation Centre (CASIC) International Working Paper Prize from Keele University, spoke at the international 'Planet Under Pressure' conference at the ExCeL Centre in London, and was invited to the House of Lords to participate in a High-Level Roundtable on Rio+20.

Ash adopted his current name and pronouns in March 2018 after coming out as transgender.